

**Interpersonal Psychotherapy:
Initial Casework in a Novel Standardized
Psychotherapy**

Roslyn Law

**D.Clin. Psychol.
University of Edinburgh
2000**



Table of Contents

Acknowledgements

i

Abstract

ii

1. Introduction	1
1.1 Practice Evaluation	3
1.2 Evaluation Beyond Manual Guidelines	5
1.3 Adherence and Competence	6
1.4 Practice and Outcome	8
1.5 Study Objectives	9
2. Interpersonal Psychotherapy	11
2.1 Overview of IPT	12
2.2 Phases of Treatment	12
2.3 Efficacy	19
2.4 Training	25
3. Rating Psychotherapy Adherence and Competence	27
3.1 Adherence and Competence	27
3.2 Measures	33
3.3 General and Mode Specific Competence	38
3.4 Differentiating Therapies	44
3.5 Interpersonal Psychotherapy	50
4. Competent Practice and Outcome	54
4.1 Interpersonal Psychotherapy	54
4.2 Brief Dynamic Psychotherapy	59
4.3 Cognitive Therapy	64
5. Therapist Factors Influencing Practice and Outcome	68
5.1 Therapist Characteristics	69
5.2 Therapist Experience	73
5.3 Therapeutic Preference	76
6. Patient Predictors of Outcome and Practice	81
6.1 Initial Symptoms	83
6.2 Collaborative Capacity	88
6.3 Patient Expectations	90
6.4 Personality Features	92
7. Models of Supervision	97
7.1 Methods	97
7.2 Models of Supervision	100
7.3 CBT and DBT	101
7.4 Supervision Format	103
7.5 Developmental Model	105
7.6 Process Model	108
7.7 Aim of the Study	112
7.8 Hypotheses of the Study	113

8. Methodology	114
8.1 Patients	114
8.2 Therapists	117
8.3 Supervisor	117
8.4 Training and Supervision Procedure	117
8.5 Training Manual	118
8.6 Didactic Seminar	119
8.7 Supervised Casework	119
8.8 Measures	120
8.9 Symptom Change and Outcome	126
9. Results	127
9.1 Therapist Details	127
9.2 Clinical Experience	127
9.3 Didactic Training to Supervision	128
9.4 Patient Details	130
9.5 Symptom Change	131
9.6 Group Adherence and Competence Scores	133
9.7 Competence over the Three Phases of IPT	135
9.8 Initial Sessions Adherence	135
9.9 Middle Sessions Adherence	137
9.10 TSRF	138
9.11 PRF	142
9.12 Focus Specific Strategies	143
9.13 Termination Session Adherence	145
9.14 Inter Rater Reliability	147
9.15 Competence and Outcome	150
9.16 Performance, Outcome and Therapeutic Allegiance	151
9.17 Experience and Outcome	152
9.18 Experience and Performance	154
9.19 Initial Symptom Severity and Outcome	156
9.20 Early Symptom Change and Final Outcome	157
9.21 Problems in Conducting Therapy	158
9.22 Most and Least Effective Therapists	159
10. Discussion	161
10.1 Adherence and Competence	161
10.2 Interrater Reliability	168
10.3 Competence and Outcome	171
10.4 Therapeutic Allegiance	174
10.5 Clinical Experience	177
10.6 Patient Factors	180
10.7 Competence Ratings	182
10.8 Limitations of the present study	184
10.9 Conclusion	186

Appendices	188
Appendix 1	IPT Session Strategies 188
Appendix 2	Adherence and Competence Rating Forms 191
Appendix 3	IPT UK Network Guidelines 197
Appendix 4	Therapist and Patient Detail Forms 199

Bibliography	201
---------------------	------------

List of Tables

Table 1	IPT Strategies for Initial Sessions	13
Table 2	IPT Focus Areas	15
Table 3	Final Session Strategies	19
Table 4	Integrated Developmental Model of Supervision	105
Table 6	Patients' BDI-ii Baseline, Termination and Change Scores	132
Table 7	Mean Competence Scores for Early, Middle and Termination Sessions	133
Table 8	Mean competence scores on first and subsequent casework	134
Table 9	Initial Sessions Adherence and Competence Scores: Sessions 1-4	135
Table 10	Adherence and Competence Ratings	137
Table 11	IPT Focus Selection	138
Table 12	ANOVA: TSRF Adherence and Competence by Focus	139
Table 13	Focus Groups TSRF Strategy Application	141
Table 14	Group Mean PRF Adherence and Competence Scores	143
Table 15	Role Transition Frequency and Competence Scores	144
Table 16	Disputes Frequency and Competence Scores	144
Table 17	Grief Frequency and Competence Scores	145
Table 18	Termination Sessions Frequency and Competence Scores	146
Table 19	Percentage agreement on PRF, TERF and Focus Form Adherence Scores	148
Table 20	Percentage agreement, near agreement and discrepancy on PRF, TSRF and Focus Forms Competence Scores	149
Table 21	Therapists' Experience	153
Table 22	ANOVA: Competence ratings and Training Background	155
Table 23	Mean (SD) Competence ratings by Training Background	156

List of Figures

Figure 1	Process Model of Supervision	109
Figure 2	Mean BDI-ii Baseline, Termination and Change Scores	132

WORD COUNT: 48723

ACKNOWLEDMENTS

My thanks firstly go to the therapists and patients, who completed the work that has been reviewed, and invited me to be part of the process. I repeatedly gained as much, and often more, than I offered. My deepest thanks to Dr Helen Birchall and Liz Martin who gave so generously of their own time and knowledge and were the key to getting there. Dr Chris Freeman and Dr Claire Fyvie offered invaluable support, for which I will always be grateful. Dr Jon Wilson had the idea in the first place, so I thank him for getting it all started. Finally love and thanks to JB for believing in me and seeing what I could not.

Abstract

This study examines the supervised casework of seventeen therapists, using Interpersonal Psychotherapy as a treatment for Major Depression for the first time. Adherence and competence were measured using procedures developed with reference to the treatment manual (Klerman & Weissman et al 1984). The capacity of the more diverse population of therapists now undertaking IPT training to meet adherence and competence standards was explored, as was the capacity of current supervisors to employ rating forms reliably. This study demonstrated that practising therapists, with a range of experience and theoretical influences, were reliably found to practice the procedures outlined in the Interpersonal Psychotherapy manual, with a high level of competence. Adherence levels were good in the focus area sessions, but less satisfactory during the initial and final phases of treatment. Less experienced therapists were found to be as capable of meeting training requirements as more experienced therapists, and a significant level of symptomatic relief was reported by the participating patients. Initial symptom severity did not have a detrimental effect on treatment outcome, with patients rated as severely depressed on the BDI-ii at baseline achieving recovery or clinically significant reduction in symptoms as often as patients with a moderate depression. Therapists with a psychology based training achieved a higher standard of competence than therapists trained in a psychiatry model of medicine or nursing, but the two groups could not be distinguished in terms of clinical outcome for patients. Problems in conducting therapy, reflecting potential ruptures in the therapeutic alliance were significantly related to clinical outcome and early competence.

1. INTRODUCTION

There is an ever growing demand for the use of evidence based psychotherapy (Wilson, 1995, Klerman & Weissman, 1993). This demand is generated by clinicians, eager to work within the scientist-practitioner model, researchers eager to evaluate the relative merits and weaknesses of specific interventions, health care management eager to establish that their limited resources are used as efficiently as possible (Barlow, 1994), and clients eager to know that the treatment they undertake will be to their benefit.

While still a matter of some debate, the standardisation of psychotherapy treatments in the form of manuals has been heralded as a “small revolution” in psychotherapy research (Luborsky et al, 1984, Kiesler, 1994), contributing to the empirical validation of different treatment packages. It is now common place, in conducting clinical trials of psychotherapy, to standardise the practice of specified interventions across the participating therapists (Rounsaville et al, 1986; Waltz et al 1993; Kazdin, 1994; Kiesler, 1994; Wilson, 1997). Similarly the daily practice of clinicians is influenced by the demand for a compelling evidence base in support of the interventions chosen, which will frequently result in the use of standardised and validated treatment packages.

It has been suggested that the routine use of treatment manuals, as a basis for psychotherapy research and clinical practice, may be of benefit in number of areas, including specification, evaluation and training. Luborsky (1984) offers a summary of what such standardisation can offer. It:

1. provides a basis for more objective comparisons of psychotherapies, revealing the ways in which psychotherapies are distinct from each other or overlap;
2. offers more precise measurement of the degree to which each therapist provides what is recommended in the manual;
3. facilitates improved training of therapists in the specific forms of psychotherapy.

These points indicate three different, if mutually informative, lines of interest in manualised practice. The first addresses the psychotherapy itself, allowing both comparative analyses between models, and detailed and informed exploration of the relative contribution of general and modality specific components. The second directs attention to the performance of therapists, clarifying specific expectations in practice and facilitating relevant evaluation. The final point turns to the issue of training, suggesting that by pinpointing what is to be achieved it may be possible to devise a tailored and effective route by which to arrive at this goal.

The focus of this study will be primarily directed towards the second of Luborsky's points of commendation, that of examining therapists' performance in the light of manual based recommendations. This is a complex procedure, which involves a number of stages. The first is treatment specification. Evaluation of practice may then be conceptualised in a number of ways, including adherence to recommendations, competence of delivery, and relation to treatment outcome. Interpretation of such ratings must also consider variables beyond those described in the manual, such as pre-existing therapist variables e.g. experience and therapeutic preference, and client characteristics e.g. symptomatic severity, marital status. The context of this evaluation

is also relevant, and can range from blind rating to a supervision relationship forming an integral part of the training procedure. These points indicate that while therapist' performance is the primary focus of this study, this naturally overlaps with psychotherapy specification and training procedures. Subsequent chapters will examine each of these issues in turn.

1.1 PRACTICE EVALUATION

Such an array of issues indicates that the specification of technique, while important, is only one of a number of factors relevant to therapist evaluation. The step between a printed set of guidelines and evaluation of the skilled application of its contents has received increasing attention in the literature over recent years, and has highlighted many pitfalls inherent in this task. Many different forms of psychotherapy, including Cognitive Behavioural Therapy (CBT), Interpersonal Psychotherapy (IPT), Cognitive Analytic Therapy (CAT) and psychodynamic therapies have been closely examined and defined in order to produce not just guidelines but measures of practice to guide the judgements made of therapists (Hill et al, 1992, Ogrodniczuk & Piper, 1999, Bennett, Parry & Ryle, in press). The different strategies employed in conducting supervision and practice evaluations illustrate the influence of both the models of therapy themselves and matching evaluation procedures to specific questions and training environments. Distinct models of supervision and practice evaluation will be examined in more detail in chapters three and seven.

Historically the emergence of manuals for empirically validated treatments made no small contribution to development in practice evaluation. The advances in

psychotherapy manuals from the mid 1970's onward was prompted by the assumption that the replication of good practice and rigorous evaluation in psychotherapy required greater specification of technique (Elkin Waskow, 1983). It is no surprise that the earliest manuals arose from the behaviourist tradition, which highly regards the clear definition of treatment techniques and procedures (Lang & Lazovik, 1963). Common ground was found with other therapy schools in research and administration requirements and the practice flourished. The support generally afforded to this approach (Luborsky & DeRubeis, 1984, Klerman et al, 1993), has meant that the scientific canon of standardisation has become enshrined in the practice of psychotherapy research. The use of therapy manuals for therapy studies is a publication requirement in many journals e.g. Cognitive Therapy and Research, and their use is evident in virtually all published studies of therapies such as cognitive behavioural therapy (Beck, et al, 1979) and interpersonal psychotherapy (Klerman et al 1984), and major psychotherapy outcome trials (NIMH, TDCRP, Elkin et al 1985, Sheffield Psychotherapy Project, Shapiro & Firth, 1987). This practice derives from a fundamental requirement of any scientific investigation, i.e. that the subject under investigation e.g. a model of psychotherapy, be clearly defined, isolated from contaminants, and observed in a reliable and valid way. In this way treatment integrity is preserved.

Evaluation of the effectiveness of therapies and therapists has immediate bearing in training and general clinical environments. Specific practices and models may be included in pre and post qualification training programmes and continuing professional development plans as a result of their evidence base, although the specific influence of

research in this area has been questioned (Crits-Cristoph et al, 1995). This discontinuity may at least in part be due to more information being available on efficacy rather than on the ground effectiveness (Seligman, 1996). In order to meet the demands of Roth, Fonagy & Parry (1996), to evaluate whether research findings result in the right thing being done in general practice, that being done well and resulting in the right outcome, specification and evaluation continue to be crucial e.g. as the basis for clinical audit, service development plans etc. It is not coincidence that the single question guaranteed to be raised following completion of a didactic training in psychotherapy skills is "what supervision is available?", as therapists conduct their own personal audit and anticipate funding discussions with managers. The limited realism and generalizability of effectiveness evaluations cannot be assumed to be an adequate excuse for not employing the guidelines and monitors which are a component part of manualised and evidence based practice, and which contribute to the delivery of accountable practice. Neither will research and theory move forward in productive way if informed feedback from clinicians, on the shortfalls of proposed practice, is not provided.

1.2 EVALUATION BEYOND MANUAL GUIDELINES

While the availability of psychotherapy manuals is argued to offer some benefit in most cases, closer examination highlights a vast range in content, extending from general guidelines to detailed prescription for practice. Consequently it is unsafe to attempt to draw a single conclusion about the use of manuals, but rather each should be evaluated on its capacity to facilitate the, comparative, practice and training procedures highlighted by Lurborsky (1984).

In addition it is important, if analyses are to be meaningful, to consider the factors which have bearing on therapists' ability to follow the recommendations. Major psychotherapy investigations, such as the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP), have attempted to do this by giving attention to the relation between mode specific and general skill (Rounsaville et al, 1987), therapeutic alliance (Krupnick et al, 1994), and patient difficulty and therapist performance (Foley et al, 1987). Similarly the Sheffield psychotherapy studies explored differential therapist performance (Shapiro et al, 1989) and the impact of symptom severity (Shapiro et al, 1994).

An additional factor to be considered is the context of the training e.g. in preparation for participation in a psychotherapy research trial or as part of ongoing clinical training and development. Such contextual factors may highlight differences in therapist experience and therapeutic allegiance, which have been shown to be influential in practice evaluations (Robinson, 1990). The contribution of therapist and client factors to process and outcome evaluations will be discussed in more detail in chapters five and six.

1.3 ADHERENCE AND COMPETENCE

Psychotherapy investigations rely on appropriate assessment measures to evaluate whether and how well the intended therapies are implemented. The existence of a manual does not ensure that its directions are followed, thus a process of evaluation based on the manual's prescription can both clarify the basis on which therapists are

being rated, and potentially enhance outcome, if the evaluation is used to modify practice to bring it closer to that prescribed (Frank et al, 1991), although this latter suggestion is hotly debated in the face of conflicting results. Prior to the emergence of treatment manuals the criteria by which therapists were judged were much more loosely defined, and consequently practice was evaluated on a more general level, if at all. Moncher & Prinz's (1991) review of outcome studies identified a specific treatment protocol in only 26% of trials, with only 13% presenting data on therapist competence.

Wilson (1997) suggests that adherence to a manualised treatment is likely to facilitate more focused treatment, and assist therapists in learning specific treatment strategies and skills, thus widening their skills repertoire rather than narrowing it. It may be argued that therapists, especially novices, are reassured by having a structure to their practice and a contiguity of theme throughout the intervention, especially when this has been demonstrated to be effective in controlled outcome studies. For more experienced therapists the opportunity to clarify their interventions and target specific areas of practice for additional attention, as well as having reference material for teaching and supervision may be of considerable benefit (Kiesler, 1994).

The impact mode specific adherence may have on general skills and responsivity is however a concern for authors who promote clinical eclecticism and individual case formulation (Persons, 1991), and fear that adherence to a manual will "obfuscate clinical artistry" (Davidson & Lazarus 1995). Henry et al (1993) concluded that mode specific training does not always lead to good outcome in clinical practice, and reported

that when experienced therapists were trained to perform a manualised therapy it actually produced negative effects on practice, which they interpreted as a difficulty in assimilating an externally imposed formulation which did not match therapists established patterns of thinking and perception.

These apparently inconsistent findings highlight the importance of competence ratings in addition to measuring adherence to guidelines. Skilled delivery of psychotherapy is more than the sum of its parts. The timing and sensitivity of interventions, consideration of the stages of the therapy, and evaluation of the client's responsiveness and motivation are but a few of the factors which are very difficult to manualise, but which will contribute to the quality of the therapy exchange. In order to make such an evaluation the original question of whether the therapy was delivered must be supplemented with an additional rating of how well this was done (Waltz et al 1993). Chapter three will review current literature on psychotherapy adherence and competence ratings.

1.4 PRACTICE AND OUTCOME

Wampold (1997) argued that standardisation of practice may be self defeating, focusing attention on between group differences, which has often offered fallow ground for investigating outcome differences, and by design precludes the examination of wider contextual and therapist led differences which may be of greater benefit and interest. Klein (1997) also argues that manualised treatment has failed to demonstrate its necessity by virtue of the fact that trials using manualised treatments have not shown increased treatment efficacy or outcome variance. Consequently he argues that

the time consuming and financially demanding practice of using manuals to distinguish different forms of treatment “expensively addresses a non-existent problem” (p929). As already noted the remit of treatment manuals extends beyond that suggested by Wampold (1997), and much of the literature examining standardised treatments addresses the very issues which they are supposed to preclude e.g. the interpersonal focus of the individual’s presenting difficulties (Wolfson et al, 1997), timing of the intervention (Kupfer et al, 1989) and the impact of comorbidity (Feske et al, 1998). In addition studies of standardised therapies are not devoid of significant finding e.g. IPT has been repeatedly demonstrated to produce superior results to an unmanualized package of supportive psychotherapy (Markowitz, 1995; 1998). The relative value of having an empirically validated treatment package, and being able to operate with clinical flexibility in response to ongoing assessment of the client’s needs and capacity, is often debated with reference to the impact this has on clinical change, and this will be discussed in more detail in chapter four.

1.5 STUDY OBJECTIVES

The present study aims to examine the training, and more specifically supervision process, undertaken with a number of therapists learning to conduct Interpersonal Psychotherapy (Klerman et al, 1984). Details on the development and structure of the Interpersonal Psychotherapy model are presented in chapter two. The specific purpose of the training differed for the individual therapists in the study. All therapists were learning to use IPT in order to introduce this psychotherapy model as a novel intervention in clinical practice in the NHS. Some therapists were preparing to participate in a randomised controlled study of IPT and Cognitive Behavioural Therapy

(CBT) (Beck, 1979) as treatments for anxiety and depressive disorders in primary care setting. The introduction of IPT to the UK is a relatively new innovation, and consequently therapists had to be trained in the use of this model. This study will focus exclusively on the IPT arm of the training and supervision process. Other participants were being trained with a view to developing service and training provision in their local area.

2. Interpersonal Psychotherapy

IPT began life as a psychotherapy control condition in a research trial designed to test the comparative efficacy of different treatments for the prevention of relapse in a group of adult outpatients with recurrent depression (New Haven- Boston Collaborative Depression Research Project, 1968). By specifying the different components of the psychotherapy condition it was hoped that this would achieve a standardisation of practice within this condition.

“Our original intent was not to develop a new psychotherapy, but to describe what we believed was reasonable and current practice with depressed patients and might fall under the rubric of short term supportive psychotherapy” (Klerman & Weissman, 1993a, p4).

The success of this novel treatment intervention (Klerman et al 1974; Weissman et al 1974), and its subsequent success as an acute intervention strategy (DiMascio et al 1979; Weissman et al , 1979), with treatment specific improvement on social functioning at one year follow up (Weissman et al, 1981), led to the treatment being operationalised and published in a treatment manual (Klerman, Weissman, Rounsaville & Chevron, 1984; Weissman, Markowitz & Klerman, 2000), which has served as the basis for much empirical study.

2.1 OVERVIEW OF IPT

IPT in its original form, is a time limited, present-oriented, weekly, individual outpatient treatment for depressed patients (Klerman et al, 1984). The interpersonal rationale behind this therapy draws on a wide range of literature. This includes Meyer's (1957) view of mental illness as an attempt to adapt to changes in one's environment, Sullivan's (1957) definition of psychiatry as the field of interpersonal relations, Bowlby's (1969) work on attachment, and the many studies examining the interpersonal consequences of depression and protective potential within the interpersonal network e.g. Brown, Harris & Copeland, (1977). IPT readily acknowledges, the role of genetics, biochemical, developmental, and personality factors in the causation of and vulnerability to depression (Klerman et al, 1984), while focusing on the connection between the onset of depressive symptoms and current interpersonal problems as a pragmatic treatment focus.

2.2 PHASES OF TREATMENT

IPT breaks down the standard sixteen week intervention into three related stages. In so doing it specifies particular goals for each stage and proposes a range of strategies by which these may be achieved. The first stage corresponds with the assessment needs of any new treatment and spans the first 3-4 sessions (Table 1). The tasks for this phase include taking a thorough psychiatric history, making an explicit diagnostic evaluation with reference to recognised criteria i.e. DSM-IV or ICD-10, engaging the patient in the sick role (Parsons, 1951), which brings with it responsibility to work towards recovery, conducting a detailed review of the patient's interpersonal resources

and setting, and establishing an interpersonal focus for treatment, based on the interconnections apparent between the other factors. Particular attention is given to

Table 1. IPT: Strategies for the Initial Sessions

A.	Dealing with the Depression
	<ol style="list-style-type: none">1. Review depressive symptoms2. Give the symptoms a name.3. Explain depression and the treatment.4. Give the patient the "sick role".5. Evaluate the need for medication.
B.	Relate Depression to Interpersonal Context
	<ol style="list-style-type: none">1. Review current and past interpersonal relationships as they relate to current depressive symptoms. Determine with the patient the:<ol style="list-style-type: none">a. nature of interaction with significant persons;b. expectations of patient and significant persons from each other and whether these were fulfilled;c. satisfying and dissatisfying aspects of the relationship;d. changes the patient wants in the relationships.
C.	Identification of Major Problem Areas
	<ol style="list-style-type: none">1. Determine the problem area related to current depression and set the treatment goals;2. Determine which relationship or aspect of a relationship is related to the depression and what might change in it.
D.	Explain the IPT Concepts and Contract
	<ol style="list-style-type: none">1. Outline your understanding of the problem.2. Agree on treatment goals (which problems area will be the focus).3. Describe procedures of IPT: "here and now" focus, need for patient to discuss important concerns, review of current interpersonal relationships, discussion of practical aspects of treatment - frequency, times length, missed appointments.

Klerman et al, 1984

interpersonal changes occurring proximal to the onset of symptoms to establish this focus. IPT is designed to focus on current interpersonal issues but this initial phase also considers the historical context of the current picture, exploring any repetitive and dysfunctional patterns, which may have emerged over time in the person's relationships. While this does help to focus attention on the current manifestations of interpersonal difficulty, which may be amenable to change, it does mean that IPT is less well placed to address longer standing interpersonal themes and style, which often have significant bearing on current symptoms and social functioning.

The use of medication is assessed during this phase and decisions are made based on symptom severity, historical response to treatment and patient preference. Patients can be seen for IPT alone or in combination with a medical intervention, both having been shown to have good outcome, although the combined treatment has generally demonstrated a superior response (DiMascio, 1979, Elkin et al, 1989, Karasu, 1990). However caution should be shown in the light of the detrimental impact demonstrated when IPT has been combined with a pill placebo (Frank et al, 1990). This may reflect a conflict in formulation or attribution, with consequences for treatment engagement and collaboration i.e. if a patient expects the medication primarily to effect a change he/she may engage less enthusiastically in the work of psychotherapy. If there is no medication effect then the psychotherapy is consequently also vulnerable to poorer outcome. If this is the case it would have implications for those patients receiving combined treatment, but experiencing no discernible benefit from the medication, if a strongly collaborative alliance with shared interpersonal goals had not been established and safeguarded.

Table 2. IPT Focus Areas

Interpersonal Role Transition

Depression associated with role transition occurs when a person has difficulty coping with life changes. In these cases the transition is more likely to be experienced as a loss by the person.

Grief

IPT deals with depression associated with abnormal grief reactions, which result from failure to go through the various phases of the normal mourning process.

Interpersonal Role Disputes

An interpersonal role dispute occurs when the patient and at least one other significant person in their life have non-reciprocal expectations about their relationship. In these cases the dispute is usually stalled or repetitious, with little immediate hope for improvement.

Interpersonal Deficits

Interpersonal deficits are chosen as the focus when a patient presents with a history of social impoverishment which involves inadequate or dissatisfying interpersonal relationships.

Diagnosis is taken as an opportunity to educate the patient about depression, acknowledging that it is a common, often episodic disorder, which has been extensively studied. Information is also presented about IPT, providing a rationale for this method of treatment and information about the demonstrated efficacy of this approach. This

positive presentation is used to combat the despair that many depressed patients feel about their situation and to promote hope in a positive prognosis. This can be very reassuring for many clients but may also be a complex task outside of clinical trials, where most patients present with multiple diagnoses which may significantly influence their scope to gain from an interpersonal approach (Frank et al, 2000, Brown et al 1996).

The different strands of the assessment are drawn together to explicitly link the depressive symptoms to a central difficulty within the patient's interpersonal situation in a focused formulation, which will form the basis of the second stage of treatment. IPT is designed to address four basic problem areas, and again the model defines the goals and strategies for this stage of the intervention (Table 2).

The focus area is explored in terms of the individual's own experience, tailoring the formulation and application of individual techniques to the personal history and current circumstance. In this way focus choice is not an alternative to formulation but a guide for it. Each of the focus areas has specific goals laid down, and strategies by which to achieve these goals are provided. During the second phase of therapy, the middle sessions, which run from sessions four to twelve, these strategies are implemented according to the focus choice.

The value of a single focus is worthy of further consideration. This strategy is often identified by therapists new to IPT as being at odds with routine clinical practice, which may attempt to be more flexible in addressing the multiple current concerns identified

when patients present for treatment. This may reflect the therapist's anticipated difficulty in maintaining a theme throughout treatment, even when it has been evaluated as central to the current complaint, and so potentially identifies vulnerabilities in both therapists and model. Wolfson et al (1997) have presented some preliminary evidence to suggest that the use of dual foci is not detrimental to treatment outcome. Fairburn (personal communication) has also argued for the more flexible application of the focus models, preferring to draw themes from the current interpersonal environment, which may extend over more than one of the focus areas. Although not specifically addressed in these publications, review of his published IPT research trials would suggest that this approach is not disadvantageous to patients (Fairburn et al 1991, 1993, 1995; Agras et al, 2000).

During the middle phase the patient is focused on recent interpersonal events and symptomatic experience, which the therapist helps the patient to link in the context of the focus selected. IPT therapists take an active, supportive and hopeful stance to combat the demoralisation associated with depression (Klerman et al, 1984). The patient and therapist's respective roles change and develop over the course of the sessions. In the initial sessions the therapist takes the lead, guiding the direction of the sessions and ensuring that the necessary information is collected and adequately reviewed to allow the different tasks to be completed. Having selected the focus, the respective responsibilities of therapist and patient are reviewed and modified. The dyad now shares a formulation of the patient's difficulties and it becomes the patient's responsibility to review their weekly experiences for instances that may be fruitfully examined during the session. The therapist agrees to follow the patient's lead and to

help the patient maintain the focus originally negotiated. The therapist also has responsibility to apply the prescribed strategies to facilitate this review of current interpersonal experiences. The therapist emphasises the options for change which exist in the patient's life, which the patient may have been prevented from seeing or pursuing by the experience of depression. IPT is a therapy that aims to motivate and facilitate change, and consistently works towards this goal. This change is defined in terms both of the depressive symptoms, the aim being to reduce them, and the interpersonal context, the aim being to facilitate positive functioning and resolution of difficulties.

Crits-Christoph (1998) reported that the treatment focus in IPT was less flexibly implemented than in CBT, and this may reflect the underlying assumptions of an IPT intervention i.e. that the current depression should be the target of treatment, and that this occurs in an interpersonal context. While this is undoubtedly valid for many clients, it does not always reflect either the clients' conceptualisation of their problems or their aims in engaging in therapy. Some clients may be amenable to the formulation and draw benefit from the dual symptom and interpersonal focus, particularly those who previously lacked an explanation or were initially warding off discussion on clearly difficult interpersonal dynamics. Others however may have multiple pressing demands which are very difficult to distil into a narrower meaningful focus, and interpersonal issues may be at best secondary and at worst largely irrelevant to their primary difficulties and concerns, or chronic and intractable with minimal scope for productive short term gain. In such cases it is unlikely that IPT will be of maximum benefit and guidance on other forms of treatment may be appropriate. While it would be unrealistic

to imagine that a single model could provide treatment for all, the fundamental assumptions of IPT may be seen to impose limitations on its application along with client and therapists resistance, to meeting the challenge of a focus and organised approach to problem management and resolution.

The final phase of IPT runs over the last four sessions, and operates in conjunction with the completion of the middle sessions work. In this phase the issues related to termination of therapy are directly addressed (Table 3). There are multiple aims for this phase of therapy, including review and consolidation of therapeutic gains, acknowledgement of the loss of the therapeutic relationship and preparation for independent functioning and methods of addressing re-emergence of depressive symptoms in the future (Appendix 1).

Table 3: Final Session Strategies

Termination

1. Explicit discussion of termination.
2. Acknowledgement that termination is a time of grieving.
3. Move toward patient recognition of independent competence.

2.3 EFFICACY

Efficacy data on this standardised psychotherapy has been widely published and reviewed (Jarrett & Rush, 1994, Klerman et al 1993; Weissman & Markowitz, 1994;

Frank & Spanier, 1995). The largest scale evaluation was conducted in the NIMH Treatment of Depression Collaborative Research Program (TDCRP) (Elkin et al, 1985). This produced numerous publications reporting on various aspect of IPT's performance as a treatment for depression, and the factors influencing its efficacy. In the primary analysis, beneficial effects were demonstrated for all four treatments - IPT, CBT, Imipramine plus clinical management (unstructured supportive encouragement) (IMI-CM) and placebo plus clinical management (PLA-CM), in this multi-site trial (Elkin et al, 1989). At termination significant reduction in depressive symptoms and improvement in function was demonstrated for all groups, with few statistically significant differences between them. When response was evaluated in terms of patients achieving the predetermined recovery criteria, six or less on the Hamilton Rating Scale for Depression (HRSD, Hamilton, 1967) or nine or less on the Beck Depression Inventory (BDI, Beck, 1961), IPT demonstrated a significantly superior performance to placebo for the intention to treat sample, and in those patients who completed at least twelve sessions. 55% versus 29% reported minimal or no symptoms for at least eight weeks after treatment was completed. Those patients who recovered with IPT were reported to remain symptoms free for a mean duration of 67 weeks over the 72 weeks follow up. While these results were encouraging for psychotherapy at the time, it should not be overlooked that even in the more successful completer sample, close to half of the IPT patients were not symptom free at the end of treatment, and approximately one third of those who recovered relapsed during the eighteen month follow up.

When initial severity of depressive symptoms was used to divide the original sample, IPT demonstrated a differential treatment effect in the more severely depressed group,

producing significantly superior results to the PLA+CM group (Elkin et al, 1995). Surprisingly no differential effect was found for IPT on a measure of social functioning (Imber, 1991), as had previously been reported (Weissman, 1981), but this may reflect the timing of the assessment, as Weissman (1981) did not report a significant effect until one year after treatment, while the NIMH study reviewed change at the termination of treatment. IPT was found to be most efficacious for those patients experiencing lower levels of social dysfunction (Sotsky et al, 1991), perhaps indicating the skills base this model of therapy makes use of in addressing depressive symptoms. The objective of IPT is to return individuals to their premorbid functioning level, consequently the better the relationships they had prior to the depression the better able they are to make use of this resource in overcoming their symptomatic distress. This point may be illustrated by the finding reported for those patients with a co-morbid diagnosis of personality disorder in this sample. Shea et al (1990) examined outcome for this sample and found that although there were no significant differences in mean termination HRSD scores for patients with and without a personality disorder diagnosis, significantly fewer with a comorbid personality disorder diagnosis met recovery criteria, and they reported a poorer outcome in terms of social adjustment. The authors suggested that although there was some gain in terms of relationship functioning, the level of gain may have reflected this sample's premorbid functioning level i.e. they were returned to their previous social functioning level which was lower than the sample without a personality disorder. The limitations of the IPT model for patients experiencing longer standing interpersonal difficulties are reflected in the authors' recommendation that the deficits focus i.e. specific attention to long standing social isolation and dissatisfaction, be used only infrequently, and recent work on adapting

IPT for patients with a dysthymic disorder (Markowitz, 1998) or social phobia (Weissman et al, 2000) is recommended to enhance the performance of the IPT model for this group of more chronically distressed patients.

The difficulties experienced by patients with longer standing interpersonal difficulties received minimal attention in the development of IPT, and is only latterly coming to the fore for authors interested in different diagnostic groups e.g. social phobia (Lipsitz et al, 1999). These were not the patients for whom IPT was originally conceived, and the need for more attention to the self concept which was absent in the original model is more apparent in pilot work with patients with difficulties which are less clearly acute and episodic. This work awaits clear empirical validation in the context of IPT casework.

The aim and design of this NIMH TDCRP illustrates the support afforded the use of manualised therapies by the authors. The two main goals were stated to be: 1) testing the feasibility and usefulness of the collaborative clinical trial model in the area of psychotherapy research and 2) within the context of this model, to test the effectiveness of two brief psychotherapies for the treatment of out-patient depression (Elkin, 1994). The simultaneous replication of a single study design across different sites was proposed as a means of providing more substantial evidence of the consistency and generalizability of the findings, and to contribute directly to the study of the efficacy of specific forms of therapy for specific patient groups, an area which had been inadequately addressed in the literature until that point. The authors considered it a necessary condition of the study that the different treatments were carefully defined

and their delivery carefully monitored in order to confirm that they were conducted as described, and so allow them to meet the specific aims of the evaluation. To this end experienced therapists were carefully selected to undertake training programmes specifically designed by the treatment authors and based on the treatment manuals (Chevron et al, 1983; Rounsaville et al, 1986), and preliminary case supervision was monitored with reference to specially designed measures of competence derived from the original manuals (Appendix 2). Only therapists who met predetermined competence criteria were allowed to participate in the actual study, and competence was consistently monitored throughout their participation. Following this procedure the authors reported that the different treatment interventions were discriminated readily (Hill et al, 1992) thus facilitating their comparative evaluation, and so making a positive contribution to the development of psychotherapy research.

Since its inception as a treatment for adults with depression, IPT has been used with a number of different populations, offering further evidence of its efficacy. In a recent review Markowitz (1998) reported work being conducted on IPT as a maintenance therapy for successfully treated recurrent depression (Frank, 1991), with depressed older adults (Reynolds et al, 1999), with depressed adolescents (Mufson et al, 1991;1993;1994; & 1999), with depressed HIV+ patients (Markowitz, 1995;1998), with depressed primary care patients (Schulberg & Scott, 1991, Schulberg et al, 1993), as a conjoint therapy for depressed patients with marital disputes (Klerman & Weissman, 1993), as an alternative to medication in antepartum and postpartum depression (Stuart et al. 1995, Spinelli, 1997, O'Hara et al, 2000), and as an acute and maintenance therapy for dysthymic disorder (Markowitz, 1993). Positive results have

also been reported on the use of IPT as treatment for Bulimia Nervosa in individual (Fairburn et al, 1991, 1993, 1995, Agras et al, 1999) and group format (Wilfey et al, 1993, 1999), and social phobia in individual (Lipsitz et al, 1999) and group format (Weissman & Jacobson, unpublished). Pilot work is also being conducted on IPT for anorexia nervosa (McIntosh et al, 2000), panic disorder (Arzt, van Rijsoort, unpublished), post traumatic stress disorder (Krupnick et al, unpublished; Law, unpublished), body dysmorphic disorder (Veale et al, unpublished), somatization disorder (Stuart et al, unpublished), post myocardial infarction depression (Stuart & Cole, 1996), depressed patients with physical disabilities (McAnanama & Gillies, unpublished), primary insomnia (Muller-Popkes & Hajak, 1996), and borderline personality disorder (Dawson, 1988, Marziali & Munroe-Blum, 1994). Treatment manuals have been produced for many of these adaptations to allow replication and continuity of clinical practice. As with other treatment forms designed originally with reference to one patient population e.g. cognitive-behaviour therapy for depressed patients, IPT has developed to widen its treatment remit.

As is evident in the preceding review the bias of attention in the IPT literature has directed towards treatment outcome studies. This is clearly not without exception e.g. Ablon & Jones (1999), Krupnick et al (1994), Rounsaville et al (1987), but it has meant that the mechanisms of change have not been sufficiently illuminated through the work completed to date. As IPT is more widely disseminated it has become increasingly apparent that the recommendations do not match common practice in many settings e.g. the number and frequency of appointments is double that of typical primary care services in Scotland, yet there is no empirical demonstration of the optimal package.

Preliminary evidence has been produced supporting the use of shorter interventions in primary care (Graham, personal communication, Swartz, personal communication), and as the model is more widely applied more rigorous evaluation of the active treatment components and influential boundaries will be both informative and important.

2.4 TRAINING

The academic origins of IPT as a research condition gave rise to particular training practices to allow therapists to be included within the confines of a research protocol. It has only latterly been more widely disseminated among clinicians and in clinical psychology and psychiatry training programs. Requests for training have significantly increased, particularly in the UK recently, and the availability of training programmes has demonstrated a sharp increase. This leaves the trainers grappling with the question of the most efficient and effective ways to deliver training and supervision, and on a more pragmatic level, the amount of training and supervision necessary to ensure the dissemination of a consistent model of IPT. Guidelines for training in IPT have been produced by a newly established network of IPT therapists and trainers to support and facilitate the dissemination of the model (Appendix 4), and these guidelines are currently under review by the International Society for IPT (ISIPT). This increase in interest has been fuelled by the increasing body of literature reporting IPT's efficacy, its application with different groups of depressed patients and with patients with non mood related disorders, and its inclusion in reviews of efficacious antidepressant treatments (Robinson, 1990, Karasu, 1990). Practice guidelines have been published for mental health professionals (Karasu 1993) and primary care practitioners (Depression Guideline Panel, 1993) and each included IPT as an acute and maintenance treatment

for depression, used alone and in combination with medication. Focus on evidenced based therapies (Wilson, 1997) and economic pressures (Sturn, 1995, Krupnick & Pincus, 1992) have also promoted growing interest in defined, proven and time limited treatments. IPT may be seen then as a good example of one of the manualised treatments, employed both in the research setting and in general clinical practice, which has been given attention in the literature.

3. RATING PSYCHOTHERAPY ADHERENCE AND COMPETENCE

3.1 ADHERENCE AND COMPETENCE

Treatment manuals moved beyond more general psychotherapy texts in providing guidelines for therapists to follow in conducting therapy (Svartberg, 1989a). This involved the explanation of rationale, strategy and technique. Such explanation does not however ensure the delivery of the said treatment, and measures of therapy adherence were developed to monitor treatment integrity.

The objective of such measures is to provide a concrete basis for evaluation of therapists' performance, allowing comparison of practice with ideal. Adherence measures specify the interventions therapists must complete if they are to be regarded as having delivered the intended treatment e.g. symptom review and link to interpersonal focus in IPT, delivery of a written reformulation letter in Cognitive Analytic Therapy (CAT, Ryle, 1991). Some measures will also provide a measure of general psychotherapy skill, distinct from mode specific interventions, and others will provide details of proscribed interventions, regarded as specific to other psychotherapy models and a deviation from intended practice e.g. Hollon et al (1988). The content of the individual scales is determined by the question they are designed to address e.g. is the therapist performing the intended therapy e.g. Cognitive Therapy Scale (CTS, Young & Beck, 1980), what is the relative contribution of different therapy styles e.g. Interpretive and Supportive Technique Scale (ISTS, Orgadniczuk & Piper, 1999), or can two models of psychotherapy be distinguished in practice e.g. Minnesota Therapy Rating Scale (MTRS, DeRubeis et al, 1982)?

The potential benefit of such evaluations has been further extended with the inclusion of the concept of competence, although this is not a universal component of practice evaluation scales (Waltz et al., 1993). Competence considers the skillfulness of the intervention employed, moving beyond simple concepts of absence or presence to more complicated contextual concepts such as appropriate timing, responsiveness and completeness. Such measures address issues more closely related to the quality of the intervention, although the purpose of individual measures should always be reviewed to gauge how this is done e.g. the Collaborative Study Psychotherapy Rating Scale (CSPRS, Hollon et al, 1988) specifically distinguishes completeness from quality, and measures the former. Nonetheless this multiply defined review of skill and quality adds an important dimension to the information provided by adherence measures. To illustrate, consider the scores for a technically consistent therapist who ensured adherence by reading therapeutic interventions to the patient from the manual. It may be difficult to fault the individual on choice of intervention but the flow, flexibility and responsiveness of the interventions would be unlikely to meet requirements. Provision for competence ratings is therefore an important consideration in evaluating the utility of such measures.

Snyder & Wills (1991) however have argued that the assessment of competence is unnecessary when experienced therapists are used i.e. competence is assumed. This position is potentially vulnerable as the specific relation between experience and competence has not been widely researched, often being overshadowed by research examining the link between experience or competence and outcome e.g. Hattie et al

(1984). In addition, the focus of competence judgements may vary, e.g. generic psychotherapy competence versus mode specific competence, and experience may be defined in variety of ways e.g. general experience or with the treatment model or specific client groups. With such a multiplicity of concepts it would be important to clearly specify what was being assumed. The impact of experience will be discussed in more detail in chapter four.

One report which addresses some aspects of this issue is Rounsaville et al (1986), which examined the capacity of therapists with different levels of general experience, defined by years of practice, to maintain competent adherence to a new model of therapy - IPT. The study examined the impact of training on two groups of therapists with different initial levels of experience, and followed the course of their skill acquisition. The first group were highly experienced (M=15 years) psychiatrists and psychologists with a psychodynamic background. They were found to perform very well on their first cases using manual based treatment. The second group were reported to have less general experience (M=6 years) when starting their IPT casework, and unfortunately their previous psychotherapy training was not reported. They were rated as less competent on their first cases than the former group, but showed gains in performance over subsequent cases.

It is of note that the experience which is reported for the more highly competent group involves psychodynamic training and practice, which is consistent in its theoretical origins with the IPT model. The greater ease with which these therapists adopted and applied the model may reflect technical and theoretical continuity between standard

and novel practice in addition to, and perhaps even beyond, greater duration of general experience. Thus theoretically consistent experience and competent novel practice may produce a higher correlation than diverse experience. It is yet to be demonstrated that experienced therapists from different theoretical orientations would perform in a similar manner following training in IPT, and Henry et al (1993) have suggested that such theoretical shifts may have a poor impact on some dimensions of competence. In addition it is noteworthy that the second group of therapists were rated as less competent on initial casework but not incompetent, which raises the question, how competent is competent enough? According to Rounsaville et al (1986) all therapists were performing at an "acceptable" level on their initial casework, which would suggest that experience is less important in predicting which therapists will achieve a minimum level of competence with training. Henry et al (1993) reported a similar finding in terms of adherence when independent raters evaluated time limited dynamic psychotherapy sessions conducted by senior staff and those of therapists in training.

Waltz et al (1993) discuss the use of competency ratings in detail and define therapists' competence as the extent to which relevant contextual factors are noted and accommodated in treatment delivery. Factors to be considered include stage of therapy, degree of impairment and comorbidity, among others. Waltz (1993) argues that such assessments enhance the quality of the evaluation being conducted in a number of ways e.g. establishing differential competence between therapists (Kingdon et al, 1996), and site differences in multi-site trials. Such information is also of use in providing feedback for improving the original manual to facilitate exegesis of the model, and for training purposes, consistent with Luborsky's (1984) goals of standardization. In

the latter example this allows supervisors to comment on the differential competence with which specific techniques within a treatment package are delivered by individual therapists, and so facilitates the identification of therapists who require additional training and supervisory support. Thus the exercise does not simply evaluate a treatment intervention as a whole but the individual components as well. Such detailed assessment of performance may also be of benefit to the therapist, directing attention to specific areas where performance may be improved or changed, and so potentially resolving some of the mystery surrounding treatment failures or difficulties. Such feedback does assume a collaborative process of supervision and evaluation however, and may be less viable as a goal if blind rating by external examiners is the sole rating system employed. Again the utility of individual scales is determined by their match with the goals to be accomplished.

There are a number of practical considerations to making and evaluating this kind of therapist assessment. Waltz et al (1993) argue the importance of considering context i.e. what constitutes competence in one mode of therapy would not necessarily be rated similarly in another, therefore there are no universal evaluations of competence. Context is also important within individual therapies, as the development of a treatment intervention over time will introduce new tasks and strategies which often cannot simply be interchanged e.g. directive questioning which ensures a comprehensive assessment may be appropriately replaced with a higher proportion of exploratory interventions in a specific and agreed area later in the course of treatment. In addition, while treatment manuals offer guidance on the specific techniques to be employed they are not similarly rigorous in specifying the criteria by which their implementation may be rated as

competent or adherent. Consequently we are left to debate the relative value of comparisons with “an established level of average competence (Shaw, 1984), and expert’s sense of what competence is (Shaw & Dobson, 1988), or the fit of the intervention and client problem or case formulation (Silberschatz, et al 1985). Waltz et al (1993) note that the measures generated from different manuals vary in their complexity and specificity and consequently in the expertise needed to use them. Different sources of information may be used to rate the therapist’s performance e.g. recordings or verbal description of process, and the unit of analysis chosen may also vary e.g. segments of or complete sessions. Such variations raise a number of practical questions and may determine the utility of a given measure i.e. the application of a measure which requires experienced therapists to review entire sessions has very different resource implications from a measure which employs graduate students to review segments of tape. It would seem then that the potential for different manualised forms of therapy and their methods of monitoring, to enhance a therapist’s clinical performance is not uniform, and cannot be taken as an automatic outcome of standardised practice.

Waltz et al (1993) make a number of recommendations for assessing adherence and competence:

1. All aspects of therapeutic competence should be defined relative to the treatment manual being used;
2. Investigators should carefully fit the manipulation check to the questions being asked;

3. Include unique and essential behaviours, essential but not unique behaviours, compatible but not necessary or unique, and prohibited behaviours;
4. Assess therapist competence within the therapy context.

3.2 MEASURES

A number of authors, working with a range of different therapy models, have attempted to generate standardised means to evaluate the extent to which therapists have adhered to the recommendations in treatment manuals, and also the skill they have demonstrated in delivering the treatments. The therapies under evaluation have included Interpersonal Psychotherapy (e.g. O'Malley et al 1988), Cognitive Therapy (e.g. Hill et al, 1992), Cognitive Analytic Therapy (e.g. Bennet & Parry, in press) and Psychodynamic therapies (e.g. Svartberg, 1989; Orgadniczuk & Piper, 1999).

The specificity with which adherence is defined varies across the literature, and as a consequence the different measures produced are only suitable for certain types of investigation. Orgadniczuk & Piper (1999) produced the Interpretive and Supportive Technique Scale (ISTS), which is proposed as a measure suitable for a range of interpretive and supportive forms of psychotherapy rather than being specific to one model of practice. The scale provides an indication of the relative amounts of interpretative and supportive technique employed, clarifying the predominant character of the overall intervention and measuring adherence to the respective techniques. The authors emphasise that attempts to equate greater adherence with a greater quantity of prescribed interventions is an unsatisfactory marriage of two concepts, and one rarely based on the recommendations provided in treatment manuals. In order to address

and differentiate concepts of frequency and adherence the scale is designed to produced subscale frequency scores for interpretative and supportive techniques, and a formula derived full-scale score representing the relative emphasis of techniques along a continuum. Each item is rated on a 5 point Likert scale ranging from 0 = no emphasis to 4 = major emphasis. In this way “purer” interventions are shown with more extreme full scale scores, while active but undifferentiated therapy would lie in the middle of the range.

This scale has many features in its favour. Among these are its simplicity and length. The ISTS has only fourteen items and was designed to be used by clinically inexperienced raters on review of audio recordings. Items are clearly defined and case illustrations provided to facilitate rating. The authors report high inter rater reliability on the sub scales (0.93, 0.88) and full scale (0.95), when used by graduate level raters following didactic training on psychodynamic theory, manual review and practice ratings. Such ease of application and reliability are potentially major advantages in promoting the use of a formal system of review, and point to the value of clearly specifying the observable features of technique. The value of brief and reliable means of evaluating practice cannot be underestimated in the face of almost universally limited training resources in the NHS. The provision of sub scale scores also provides specific adherence information when a comparative evaluation is not required, and the brevity of the scale would suggest that this would not be an arduous task to complete.

One weakness of the scale however is lack of attention to the quality of the intervention. The scale operates as an externally rated quantitative measure. The

predominant character of the intervention is illuminated, but the responsiveness of the therapist to e.g. different stages of therapy, are not reflected in the frequency or adherence ratings. As the authors note inexperienced raters may have an advantage in being able to follow guidelines on rating observable technique without the distraction of clinical experience and interpretation often experienced by practising clinicians who act as raters. It may be difficult however to replicate such reliability for more complex constructs and may be of limited value in facilitating a training exercise which is as interested in understanding the reasons for patterns of practice as it is in describing what they are.

The importance of such ratings is illustrated in another system of therapy rating developed for use with short-term anxiety provoking psychotherapy (STAPP, Svartberg & Stiles, 1994). The STAPP Therapist Competence Rating Form (STCRF, Svartberg, 1989a) provides a measure of both quantity and quality of interventions. Unlike the ISIT, the STCRF, like many other competence measures, is a therapy specific measure, and quality judgement demands that raters are STAPP competent therapists, unlike the more readily useable ISIT. In developing this scale Svartberg (1989) incorporated the concept of the red line cut off which indicates a serious deviation from the mean practice rating, which may be defined with reference to the individual or to a group. Shaw (1984) introduced this concept when rating cognitive therapy in the NIMH TDCRP. The cut off score can be used as a reference point both for individual sessions and for the therapy as a whole, and is of great utility to target ongoing training procedures. For example in a therapy such as IPT there are three distinct phases to the therapy with different tasks to be completed and corresponding shifts in the balance

of the working contract. Monitoring practice across the three phases would help to evaluate the therapist's skills in adjusting to and meeting the changing demands as well as general skill in adhering to the treatment model. Such a cut off also provides a very clear measure of the proportion of casework that fell within acceptable bounds, simplifying the interpretation of results.

The STCRP provides a qualitative and quantitative measure of therapist behaviour, employing a five point Likert scale, with three descriptive points and a "not applicable" rating when there is no opportunity to evaluate a given intervention. It is recommended, as with the ISIT, that the entire therapy session be reviewed to provide opportunity for the full range of appropriate interventions to be implemented. Summary and mean session scores are produced. Inter-rater reliability on individual items was reported to be in the low range when raters familiar with the therapy but new to the rating task were employed (Svartberg 1989). Acceptable inter-rater reliability (.7) was reported for the mean scores for sessions, and in a later study (Svartberg, 1999) reported good test -retest reliability ($r=.84$) with a mean interval of 2.5 years. Mean scores correlations remain low in comparison to the simpler ISIT results but are consistent with other measures of psychotherapy competence such as CBT (Vallis et al, 1986) and IPT (Chevron & Rounsaville, 1986) and this appears to reflect the greater difficulty in producing reliable complex clinical judgements. The wide variation on individual item reliability indicates that this level of assessment is not acceptable on this measure. It was noted that the most unreliable items were rarely endorsed which may suggest redundant concepts were included or training procedures may not have sufficiently promoted specific dimensions of practice in therapists or understanding in

raters. This may potentially illustrate the feedback loop that can result from such practice evaluations. One useful finding was the failure to find a significant difference between ratings based on audio and video recordings, consistent with De Rubeis et al (1982), suggesting that some flexibility in the rating procedure may be possible.

It was of note that the experience brought to the rating exercise by the different raters appeared to influence their use of the scale, apparently biasing one towards competence and one towards adherence ratings. In addition the different patterns of individual scores arriving at consistent means scores suggests that the raters reached the same conclusions for different reasons. While this may demonstrate the flexibility which is possible within a standardised model i.e. different routes to the same end, the fact that the same practice was rated differently points to the fundamental tendency of raters to interpret what they hear, even when provided with specific guidelines. This tendency highlights the importance of taking into account the nature of the experience brought by the rater, and their involvement in the training, supervision and evaluation. In Svartberg (1989) both raters were highly experienced but for one this came from clinical practice and for the other from developing the measure, providing detailed knowledge of the conceptual implication of each item. Numerous combinations of raters and measures are reported in the literature, from graduate raters blind to the psychotherapy condition to supervisors highly involved in the detail of the casework. It is likely that each combination has advantages but it is important to match needs and design to realise these, and to be similarly aware of the disadvantages e.g. involved raters potentially being influenced by a wider range of data such as their personal relationship with the therapist and/or supervision discussions which plan subsequent

interventions which could be used to supplement or in place of actual practice. Explicit definitions and guidelines for assessment should be employed in such cases, but it is important to be realistic about how much this may control, especially when measures by definition require the use of clinical judgement. Periodic comparative checks with other raters may at least provide an indication of the influence being exerted. Inexperienced raters in turn may have significant difficulty making quality judgements, and the consequent burden on experienced expert raters, who are a limited resource, may create obstacles to the wider use of evaluation procedures.

Subsequent work on the STCRP also addressed the interesting question of temporal contiguity of competence ratings (Svartberg,1999). Once again, considering the limited resources which are often available to provide detailed review of practice, it is important to consider when in the course of therapy evaluations may be most usefully made, and whether this is common to different forms of therapy. Analysis of STCRF ratings over twenty session treatments indicated minimal and non-significant change in therapists performance, and offered preliminary support to the use of an early, non-assessment session as an indicator of treatment competence. Unfortunately the small sample size again calls into question the security of this finding. Thus this interesting scale appears to have at best inconsistent psychometric validity.

3.3 GENERAL AND MODE SPECIFIC COMPETENCE

The relationship between therapist' mode specific and general competence, across different therapy models, has been interestingly addressed in a number of studies which have highlighted the potential need for more than one type of adherence and

competency measure to be used. This relationship is important because the competent delivery of individual models of therapy often implicitly assumes the competent use of general therapy skills, which are in turn regarded as inherent components rather than distinct and optional supplements. This point is however a matter of some debate for some authors (Persons, 1991).

Rounsaville et al (1987) used the Vanderbilt Psychotherapy Process Scale (VPPS, O'Malley et al, 1983) to rate dimensions of the therapy relationship, productiveness and patient and therapist activities or characteristics, and the Therapist Strategy Rating Forms to measure IPT strategies, general skills and specific techniques. While this approach provides a more multidimensional assessment it is also highly demanding on resources e.g. the VPPS has 80 items and the IPT scales involve up to 60 different ratings across the treatment. Three hundred and sixty one sessions from thirty five training cases for the NIMH TDCRP were reviewed by two of the therapists' trainers and supervisors. Inter-rater reliability was reported to be acceptable for the TSRF ($r=.71$ to $.85$ for the subscales), and acceptable but more variable for the VPPS, Pearson r s ranging from $.56$ to $.83$. This study revealed that all general therapy factors were significantly related to supervisors' ratings of IPT skills, with positive factors e.g. therapist warmth, positively correlated and negative factors e.g. therapist negative attitude, negatively correlated.

Therapist self ratings of effectiveness were also analysed, providing a comparison with supervisors' ratings, which may have been vulnerable to a halo effect as the supervisors completed both general and mode specific ratings. This second line of

rating did not alter the pattern of correlation, although the significance of some relationships was reported to be slightly weaker. Multiple regression analyses which subdivided therapists according to one of three training sites, and therapy sessions into early and late interventions again did not change the pattern of the findings. It should be noted however that with only eleven therapists the numbers are very small to justify such an analysis. In addition while inter-rater reliability scores go some way to fostering confidence in the reported results the use of two raters directly involved in the training and supervision procedure without an external judge could not control for the influence exerted by knowledge of the trainees. Therapists' self ratings provide an interesting comparison as therapists may also be vulnerable to demonstrating an averaging out effect, judging themselves to have done generally well or poorly. The positive self reports are inconsistent with other reports of therapists' discomfort with novel practice (Henry et al, 1993) and may reflect differences between models. Such therapist-supervisors comparisons have also been shown to produce inconsistent ratings, with IPT therapists' self ratings failing to correlate with supervisors ratings based on process notes or on videotape review (Rounsaville et al, 1983). Rounsaville et al's (1987) findings would however suggest that mode specific adherence and competence is not detrimental to general therapy skills nor, given the low to moderate range of variance accounted for by the significant correlations which were produced, are the concepts of general and specific adherence and competence synonymous in IPT.

The consistency of general and mode specific ratings is challenged by both Henry et al (1993) and Vallis et al (1988). Henry et al (1993) examined change in therapy

behaviour in sixteen therapists who underwent a year long training in time limited dynamic psychotherapy. Adherence was measured with the 21 item Vanderbilt Therapeutic Strategies Scale (VTSS, Butler, Henry & Strupp, 1992), which measures interviewing behaviours and specific strategies. Technical adherence is rated in terms of frequency counts, a strategy which has been criticised for holding an uncertain relation to manual guidelines, and interviewing is rated qualitatively, both employing a 5 point Likert scale. General dimensions were rated with assessment of therapeutic interaction using the VPPS (O'Malley et al, 1983) and the Structured Analysis of Social Behaviour (SASB, Benjamin, 1974). Unlike Rounsaville et al (1987), the time limited dynamic psychotherapy study examined therapists' skills before and after training rather than in relation to each other during training cases. This is important as it prevents a direct comparison of results, as does the different conceptualisations of general skills in the two studies. However the patterns of change in scores indicated a significant and positive increase in specific skills, while interviewing skills remained stable, and general skills were shown to deteriorate, although often not significantly, with therapists shown to deliver hostile and complex questions more frequently, and to spend less time evaluating the patient's feelings. An important point to consider when assessing these findings, in terms of the relationship between general and specific skills, is that the raw number of undesirable therapy interactions increased along with the significant increase in mode specific interventions, but their respective proportions did not change with training i.e. therapists were found to be doing more of most things after training. This lack of change in the proportional use of general strategies, and the stability of the interview skills which were rated for competence, would suggest that mode specific practice does not significantly impair general skills, although the

relationships between the two dimensions appears to be weaker in this evaluation than in Rounsaville et al (1987). The results of this study do seem to highlight the impact of deliberate and perhaps more self conscious practice following training. Skills which have become automatic and intuitive may temporarily be drawn into the more conscious arena along with newly acquired techniques and knowledge. This does not necessarily mean that such skills are lost or damaged but are temporarily subjected to a disproportionate level of scrutiny and editing which does not facilitate their delivery.

Henry et al (1993) speculate that this may be an effect of conducting ratings during a training period when the model is not fully internalised and is still being enacted in a deliberate, conscious way which may not characterise normal practice. Thus the first or second case a therapist conducts in a given model may be a poor indicator of their mean standards of practice. Rounsaville et al (1987) evaluated therapy conducted during the first four training cases and this additional experience may have been reflected in the results. Three other differences are important to note. The first is that the therapies under study were different in the two studies, and it may be that general skills are effected differently by different novel therapies. Secondly the therapists under study were also different, with Rounsaville et al's (1987) reporting a mean of 15 years experience, while Henry et al's (1993) sample had a mean of 5 years experience. In early IPT studies therapists were highly selected and had completed training in psychodynamic therapy and had two years post qualification practice, while Henry et al's (1993) sample has received training and supervision in psychodynamic principles and techniques but no formal training. Thus the confidence in general skill may have differed between groups. Finally the raters were also very different in the two studies

with the first employing raters very closely involved in the training and supervision process, while the second used raters who were unaware even of the training status of the cases. It seems very likely that the two groups would have very different supplementary information available, which would potentially have bearing on ratings. In practice they also had very different samples of therapy to rate, with the Rounsaville et al's (1983) raters reviewing 9-16 sessions of a treatment, while the Henry et al (1993) raters sampled only 15 minute segments of two sessions. In effect the former raters had significantly more information on both the therapist and patient which could both bias and inform their ratings. This raises the question of the suitability of using only segments of tape to rate interventions for competence.

A third study which addresses the relation between two skill dimensions, and which introduces another model of therapy, is Vallis et al's (1988) study of Cognitive Therapy conducted in the NIMH TDCRP (Elkin et al, 1985). This employed the well validated 11 item Cognitive Therapy Scale (CTS, Young & Beck, 1980) and the 28 item Matarazzo Checklist of Therapist Behaviour (MCBT, Matarazzo et al, 1965), which assesses therapists' errors in role definition, focus, and facilitation of communication. Both measures are reported to have demonstrated acceptable inter-rater reliability. An interesting variation in design in this study was the use of different raters for the different scales. CTS ratings, which were made by trainers/supervisors and external experts, were based on entire sessions, and MCBT ratings, made by a trained research assistant, were based on three five minute segments. This study yielded a non uniform pattern of correlation between cognitive therapy skills and therapist behaviours. Errors in focus and role were negatively correlated with CT competence, while errors in

communication were unexpectedly positively correlated. Detailed examination of the errors in communication revealed a specific pattern of "errors", consisting largely of brief answer questions and interruptions, which the authors argue could be seen as consistent with the active and directive style of CBT, helping patients to exit non productive ruminative cycles. This however raised the point that all general skills are not equally appropriate and prominent in different forms of therapy. The effective use of general skills may follow different patterns across different models rather than being indicative of a deterioration of one set of skills with the acquisition of another.

3.4 DIFFERENTIATING THERAPIES

Another approach to adherence and competence ratings which has found support is the construction of substantial, comprehensive measures which detail and distinguish different treatment approaches to an identified problem. Such scales may not only produce measures of general skill but also absolute and relative levels of interventions consistent with alternative techniques, and the quality or extensiveness of their implementation. As such they offer some of the same features of the ISIT i.e. predominant character of the treatment, and the STCRF i.e. competence or extensiveness and adherence, but on a much larger scale. Two examples of such scales are the 55 item Yale Adherence and Competence Scale (YACS), which measures general elements of behavioural drug abuse treatment, and critical components of clinical management, twelve step facilitation, and cognitive behaviour therapy (Carroll et al, 2000), and the 96 item Collaborative Study Psychotherapy Rating Scale (CSPRS, Hollon et al, 1988) which differentiates Interpersonal Psychotherapy (IPT), Cognitive Behaviour Therapy (CBT), Clinical Management and non specific

therapy variables in the treatment of depression. The comprehensive and consequently time consuming nature of these scales is a simultaneous strength and weakness, and significantly influences their field of application. In research trials a single measure, which accommodates the crucial components of the individual arms of an investigation, is a valuable unifying procedure reducing error variance, while in general clinical training such a range of assessment is likely to involve a measure of unnecessary detail and comparison.

One difference apparent between these two rating systems, is the format and purpose of the rating. The YACS employs a split system in which separate ratings are made for adherence and competence, while the CSPRS uses a single score to represent the extensiveness, frequency and intensity of an intervention rather than the quality specifically. In employing this split system the YACS underlines the distinct, although often related, nature of adherence and competence and moves away from global measures of competence (DeRubeis et al, 1982) to a systematic evaluation on each item. This approach appears to be justified by the reported results which demonstrated a moderate positive correlation between adherence and competence, indicating some measure of independence between the two constructs (Carroll et al, 2000). Specifically they demonstrated different levels of variability, with competence demonstrating greater stability than adherence, consistent with Svartberg (1999). The authors suggest this may indicate technical flexibility with relatively stable levels of skill, and reiterate the point that the two concepts cannot simply be alternated, and their combination may enrich the data produced. In addition distinct patterns of correlation were reported between the individual treatment specific scales and general scales, highlighting the

importance of carefully defining the nature of technical adherence and skillful competence to be measured. Carroll et al (2000) demonstrated that the individual treatments could be reliably distinguished, but that they each held different relationships with the more general assessment, goal setting and general support scales. This may illustrate why other studies have produced inconsistent results with regard to the influence of therapist relative to the therapy model i.e. different dimensions bend under different influences.

Consistent with other competence measures YACS raters are required to have both general experience of working with substance users and specific experience of at least one of the rated treatments, while the more descriptive CSPRS requires no previous experience of the modalities under review. Nonetheless both groups of raters undergo systematic training involving didactic teaching, manual review and practice ratings, evaluated with reference to acceptable standards. The YACS reports high inter rater reliability for its six scales on adherence (ICCs = .80 - .95), and competence (ICCs = .70 - .97), but some question is raised over individual items which have low inter rater reliability e.g. $r = .06$. Data for the YACS are presented on a substantial sample, 576 sessions representing 117 patients and five raters were employed, suggesting a robust evaluation of the merits of the scale.

The CSPRS (Hollon et al, 1988) was specifically designed for use in the NIMH TDCRP (Elkin et al 1984) to discriminate between the three therapies under investigation. The scale was developed in close consultation with experts of each treatment modality and with reference to the individual treatment manuals, and is the end result of six revised

versions (DeRubeis et al, 1982). Application of this measure during the pilot training phase (Hollon et al, 1988) and subsequently in the treatment phase of the study (Hill et al, 1992) demonstrated an impressive capacity to distinguish between the treatments in a reliable and consistent way across raters. Interrater reliability was reported to lie between .78 and .92 for the 28 item IPT and CBT scales, and it was of note that the specified treatments (IPT, CBT and Clinical Management) were consistently more reliably rated than the general dimensions of Facilitative Conditions ($r = .47$ to $.58$) and Explicit Directiveness ($r = .58$ to $.73$). This may reflect the greater difficulty in specifying general psychotherapy process than model specific techniques for clinically untrained raters. Importantly however the replication of Hollon et al's (1988) reliability ratings in general demonstrated the capacity for the scale to be used with consistent reliability by different raters. These results are not achieved without costs however, and like the YACS, an extensive training process is involved for raters, estimated at 50 hours for the CSPRS. The manual is reviewed in detail and didactic instruction is provided. Preliminary ratings are made and discrepancies discussed in detail. Practice rating are continued, involving 18 cases, until an acceptable level is achieved. Hollon et al (1988) demonstrated that only two raters were necessary to achieve desirable reliability ratings. An interesting dimension of the analysis was the observation that therapists' practice differs across sub scales over the course of treatment, with all therapists rating higher on the Clinical Management scale early in treatment, and later reducing for IPT and CBT therapists. This supports Waltz et al's (1993) recommendation that contextual i.e. timing, information is available to make appropriate assessment.

While it is apparent that both measures have been carefully developed and robustly tested, the demands involved in using the scales, both in training and actual rating time, seriously limit the application of the scales. The level of demand is illustrated if we appreciate that the time demand to train in using the CSPRS is equivalent to that of a therapist completing an IPT training course and conducting a supervised case. The measures were designed for specific purposes, having been developed in the course of major research evaluations, but it is difficult to see the practical utility of such extensive measures in daily practice review and supervision, particularly without an explicit competence rating from the CSPRS. This illustrates an implicit assumption that standardized competent practice is the domain of psychotherapy research, rather than the challenge which routinely faces practising clinicians, and scales are reasonably judged not only on their generalizability to other raters but also to other training and practice settings.

Wagner et al (1992) took this task forward by developing a 27 item Therapy Rating Scale to evaluate and distinguish two forms of maintenance treatment for recurrent depression – IPT-M and Medication Clinic. Although still the product of a research trial it does provide a more manageable evaluation for more general use. This scale was developed from previous attempts to distinguish IPT and CBT (DeRubeis et al, 1982), selecting the items which were most effective at distinguishing the treatments, and combining these with additional manual derived items as well as from the IPT Therapist Strategy Rating Form used in the NIMH TDCRP. These items were piloted and revised, retaining only those items with inter-rater agreement of at least $r = .85$. The final scale rated the extent to which items were used on a five point Likert scale and

reflected IPT, Medication Clinic (MC) and contaminating items (characterised as CBT or psychodynamic interventions). Results are reported on twenty four undergraduate raters employed over eight years, all of whom had to achieve .85 reliability to be included. Raters were blind to treatment condition and reviewed only a sample of tape, to avoid clues from the different session lengths. Test-retest reliability was reported at .90 when full session and sections of tapes were reviewed, but again it must be noted that the task is to identify the use of specified techniques and not to judge the quality of the therapy intervention. A minimum of five sessions were reviewed for each patient-therapist dyad ($M=14.84$, $SD=9.7$), which the authors argued was the minimum necessary to characterise the therapy relationship. Ninety-two pairs were included, reflecting the work of nine therapists. Mean inter-rater agreement on individual items was reported at .82, and treatment were correctly classified in 83% of cases. Factor analysis revealed an eight factor model which accounted for 73% of the variance, and four of these factors significantly discriminated the treatments. On the basis of these analyses it would appear that this scale had acceptable psychometric properties and can be used reliably by an extending group of raters. In addition the use of undergraduate raters using a relatively short scale, on seven minute segments of tape, has important advantages in terms of resource demands. Such a scale would have application if the task were to monitor the character of the interventions used, but is still likely to fall short if the demand were a more detailed review of the selection and quality of interventions across treatment.

3.5 INTERPERSONAL PSYCHOTHERAPY

The main work conducted to specify and then monitor adherence and competence in Interpersonal Psychotherapy was that of the TDCRP (Elkin et al, 1984). The CSPRS (Hollon et al, 1988) was devised as a discriminant measure to differentiate the different treatment arms of the study and was shown to be reliable and effective, but the IPT training and supervision process also made use of a series of evaluation forms (Rounsaville et al, 1987, 1988; Foley et al, 1987; O'Malley et al 1988). These forms specifically addressed broad interpersonal strategies (Therapist Strategy Rating Form), focus specific interventions (four IPT focus forms), early and termination sessions interventions (Early and Termination sessions forms) and general orientation and skill (Process form and Overall rating form).

The IPT treatment manual was undergoing extensive revision and expansion at the outset of the TDCRP, and as such was used as "the major instrument to define, specify and transmit the strategies and techniques of IPT "(Rounsaville et al, 1983) for both treatment delivery and evaluation. The manual offered detailed instruction about the actual conduct of the treatment, demarcated the external and internal boundaries, and charted the sequence to be followed over the different phases of the treatment. It provided an operationalized list of prescribed and proscribed techniques, detailed four general strategies for approaching the interpersonal problem depending on presentation, and provided guidelines on how to manage common problems (Rounsaville et al, 1983). Despite this extensive list of what is provided, it is also clear that much is not defined in the IPT manual. The manual is a description of the

application of therapy techniques but was not specifically intended to operate as a manual for the supervision forms that arose from it. Despite the absence of a separate rating manual there are several reports of the reliable use of the supervision forms, with acceptable standards of agreement between the raters (Rounsaville et al, 1987, 1988; Foley et al, 1987; O'Malley et al 1988). However it is important to note that these studies were largely undertaken by the originators of the model, and it is these experts who served as therapy evaluators. Subsequent reports on IPT have demonstrated a significant absence of reliability data (Lave et al, 1998; Brown et al, 1999; Blanco et al, 2001; Zlotnick et al, 2001), leaving new trainers and supervisors very uncertain on the relevance of earlier findings and the suitability of review measures which have only been shown to be reliable when used by the people responsible for designing them.

The IPT approach to ensuring a high standard of practice was to place considerable emphasis on the selection of therapists for training, and a number of criteria were set down. Therapists were at minimum :

- to be fully qualified psychiatrists or clinical psychologists,
- to have a minimum of two years post qualification experience,
- to have received training in a psychodynamically oriented framework
- to have experience of treating at least ten cases of ambulatory depression psychotherapeutically,
- to have good general clinical competence,
- and to show interest in and commitment to the IPT approach, and a lack of attachment to techniques or theories incompatible with the IPT approach (Rounsaville et al 1983).



While this may have been possible to achieve in the context of a major research trial, with the authors acting as trainers, supervisors and evaluators, this no longer describes the body of therapists being trained in IPT, and consequently has implications for training and supervision. The training, skill and knowledge assumed by the originators cannot be assumed for all trainees, or in fact for all current trainers and supervisors, and this may be reflected in early adherence and competence ratings following didactic training. This is a major gap in the IPT literature, as reports of reliability rely exclusively on data that were the result of a highly selective training programme.

When these requirements were met the rating forms were employed to conduct three levels of evaluation, including use of appropriate IPT strategies, use of appropriate IPT techniques, and monitoring the use of techniques which are not part of the IPT approach. This latter point could include non-specific but compatible techniques and proscribed interventions (Waltz et al, 1993). Supervisors and trainers have frequently been employed to rate compliance and competence in IPT sessions, and entire sessions over the course of treatment have been consistently reviewed to provide useful contextual information and to monitor the appropriate sequencing of strategies as demanded by the model. Knowledge of the whole treatment to inform evaluation of specific interventions within an individual session is a theme that emerges from the IPT literature. This may be indicative of the therapy style, which emphasises the importance of understanding the individual in the context of their social environment, influencing the model of evaluation which was developed. However in order to safeguard against the potential biasing effect of using supervisor raters, external independent evaluations

were also made of the TDCRP therapy. Acceptable to good interater reliability was reported ($r = .71$ to $.87$), suggesting that supervisors are capable of making valid and reliable judgements using the therapy manual as the guide for evaluation. Monitoring procedures following initial training are also reported to employ the red line procedure such that if two consecutive sessions fall below the cut off the therapist is offered more intensive supervision to facilitate improved performance. Once again the interplay between Luborsky's (1984) goals of standardization is apparent. This was reported to be infrequently required and readily rectified, and it would be interesting to examine whether the same pattern of adherence and competence would be observed in a less highly selected therapist sample. Rounsaville et al's (1987) comparison of specific and general dimensions of psychotherapy, demonstrated that while the general and specific rating were correlated, this was only to a modest level. Approximately sixty percent of correlations were less than or equal to $r = .35$, suggesting an independence of the concepts and a specificity to the IPT ratings. Thus ratings of IPT adherence and competence using the CSPRS and TSRF have been shown to be model specific and reliable but data has not continued to be produced to replicate these findings for the wider sample of IPT therapists and supervisors who now practice.

4. COMPETENT PRACTICE AND OUTCOME

“All have won and all must have prizes” is an often quoted conclusion on comparative psychotherapy outcome research (Luborsky et al, 1975). The fact that many studies comparing ostensibly distinct models of therapy have produced largely similar results for the therapies under study (Elkin et al, 1989; Stiles, 1986; Shapiro, 1990) has generated a number of explanations. Some have questioned the statistical validity of many studies (Kazdin & Bass, 1989), others have looked to a closer examination of the patient samples (Elkin et al, 1988), while others have highlighted the multiply defined influence of the therapists (Robinson et al, 1990). One further means of approaching this equivalence in outcome has been to look at the quality and specificity of the therapies under investigation. Given that the evaluation of practice has developed to allow detailed description of the models, and monitoring of if and how well the therapy has been delivered, then it should be possible to establish whether more internally consistent and higher quality therapy results in better outcome for patients. In this way adherence and competence evaluations would be seen to demonstrate their predictive validity (Shaw & Dobson, 1988). This question has been addressed by a number of authors, practising a range of psychotherapeutic approaches, but has failed to produce a conclusive answer, and in fact has generated directly conflicting results (O'Malley et al, 1988; Svartberg & Stiles, 1994).

4.1 INTERPERSONAL PSYCHOTHERAPY

Research conducted with IPT as the reference treatment has produced mixed results with regard to adherence, competence and outcome. One of the earliest studies

(Rounsaville et al, 1981) found that psychotherapy process variables, defined as techniques, topical focus and attendance, were not related to outcome measured by the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), and the Social Adjustment Scale-Self Report (SAS-SR; Weissman & Paykel, 1974). Individual techniques and time allocated to the major problem area were significantly correlated to end point HRSD scores, but multiple regression analysis revealed that process factors did not make a significant contribution to the explained outcome variance once the effect of prognostic patient variables and initial severity levels were taken into account.

A major vulnerability of this study was the assumption that psychotherapy process can be adequately defined in terms of time spent using a specific technique or focus. This is at best a crude measure and gives no indication of the competence of the intervention and appears to assume that more is better, which has been disputed in the competence and adherence literature and is not substantiated by manual guidelines. In addition the information on process came from therapists' reports on the Treatment Schedule form, which was completed immediately after the session. This method of rating the proportion of time allocated to each variable is potentially flawed, relying on therapists' memory of the session and judgements of time allocation and therefore potentially vulnerable to an array of distortion effects e.g. primacy, recency, effectiveness, receptiveness. Given these methodological flaws this study can at best be seen to offer uncertain support to the primacy of non-specific and patient factors, particularly pre treatment social functioning and general emotional health, in predicting outcome. These findings raise questions over the primacy of mode specific techniques in IPT, but it must be noted that these were defined in quite general terms e.g.

exploratory and reflective or directive. This did not address focus specific interventions e.g. evaluating the positive and negative aspects of the old and new role in transitions or the contribution of nonreciprocal role expectations in disputes, which may have provided a better illustration of IPT specific interventions.

In a later study on IPT O'Malley et al (1988) used data from thirty five patients treated by eleven therapists in the NIMH TDCRP. The fourth of sixteen therapy sessions was rated using the TSRF to provide a measure of competence, and therapists rated their own performance on a seven point scale of effectiveness. Outcome was measured using the HRSD, SAS, and patients' subjective ratings of change over the course of treatment on a seven point scale. Each outcome variable was assessed prior to treatment and was controlled for in the final analyses. Supervisors' ratings were compared with those of independent judges and produced adequate intraclass agreement (.60 - .80). Higher supervisors' ratings of competence were significantly related to greater patient rated change, and therapists' self ratings approached significance with this outcome measure. When therapists were split into high and low competence groups, based on supervisors' and therapists' ratings independently, the high groups were significantly different from the low groups on patient rated change. It was of note however that competence and effectiveness ratings produced non-significant, largely negative correlations with the independently rated outcome measures, and did not contribute significantly to the variance in the independent outcome measures' total scores. Competence was however found to be significant in predicting change on the apathy factor of the HRSD.

Although this study has gone some way to demonstrating a link between competence and outcome, the connection is not expansive. Patients' ratings on overall change is easily the most global outcome measure used, and the one most vulnerable to interpretation and distortion. It would seem very possible that such a rating would be influenced by factors such as satisfaction with treatment and perhaps the therapy relationship, which has been shown to be significantly related to outcome (Krupnick et al, 1994), calling into question the actual relation to competence ratings. The summary nature of the competence rating may also be important in interpreting these results. The TSRF subscales involve assessment on a number of dimensions of practice, including selection and implementation of problem oriented strategies, application of specific techniques and overall ratings of session quality. A single composite score was employed in the present study to represent competence, and it may be that this summary process masked the effect of individual components of competence on outcome.

It is of note that the two measures most directly indicative of success in treatment goals for IPT i.e. to reduce depressive symptoms and to improve social functioning were not significantly related to therapists' competence. At least for the latter measure this may be a function of timing, a factor also true of Rounsaville et al (1981). It has been repeatedly demonstrated that change in social functioning is not significantly demonstrated until some time after IPT treatment has concluded (Weissman et al, 1974; Weissman et al, 1981; Agras et al, 2000). However ratings in both studies were taken at the end of treatment, potentially missing the subsequent significant impact. The highly selected nature of the early IPT therapists must also be borne in mind when

examining these results. The therapists were selected and rated to be highly competent, and as Stiles & Shapiro (1989) warn, a null finding under such conditions cannot be assumed to indicate an absence of effect. Consistently high performance rates would not be readily detected in statistical analyses as they would generate insufficient variance. Data presented on supervisors' ratings indicate good performance levels and minimal variance ($M= 3.08$, $SD=0.6$) and the therapists self-ratings were similarly towards the more effective end of the seven point scale ($M= 4.76$, $SD=1.1$).

The clearest positive finding in the IPT literature is Frank et al's (1991) examination of the contribution of the quality of IPT sessions to the length of the well interval for patients with recurrent depression, in a three year maintenance trial. Seven minute segments of each therapy session were rated on specificity and purity of interpersonal interventions, using the Therapy Rating Scale (Wagner et al, 1992) which was derived from the IPT and medication clinic training manuals. Trained undergraduate raters blind to the treatment condition were employed, and good interrater reliability (.85) was reported. Therapy conducted by seven clinicians with thirty six patients was reviewed. Survival analyses revealed that only the interpersonal score remained significantly related to time of depressive recurrence when other explanatory variables i.e. somatic and cognitive interventions, were controlled for. This relation was further illuminated when the therapists were divided into high and low interpersonal focus groups based on the median group score. This revealed a highly significant difference in the period patients remained well, with the low interpersonal group's patients experiencing a

recurrence of symptoms after a median period of 18.1 weeks (SE 4.6 weeks) and the high interpersonal group surviving for 101.7 weeks (SE 7.7 weeks).

These studies highlight a number of the factors which are important to consider when reviewing the relation between practice and outcome. Each study purported to examine the same therapy but practice was measured differently each time, even when the core concepts were agreed upon. The dimension of practice also varied between time allocation, quality and specificity and produced different results, which highlights the potential for aspects of competent or adherent practice rather than global indices to be fruitful. Outcome was also characterised in a number of ways, again illustrating that practice can be evaluated in the light of numerous treatment goals. The basic therapy task also differed between the first two and last study, with the former aiming to get patients well and the latter aiming to keep well patients so. With such an array of factors even within one therapy it is not surprising that firm conclusions are elusive when different models are compared.

4.2 BRIEF DYNAMIC PSYCHOTHERAPY

Barber et al (1996) examined the effect of therapists' adherence and competence on patient outcome in Brief Dynamic Therapy for current major depression. Their design was similar to O'Malley et al's (1988) except that they measured the contribution of adherence and competence to predicting symptom change from the point at which techniques and skill were measured i.e. session 3, rather than from intake to control for unexplained initial symptomatic gain. The Penn Adherence-Competence Scale for Supportive-Expressive therapy (PACS-SE; Barber & Crits-Christoph, 1996) was used

by two independent clinical psychologists to rate audiotaped sessions, and outcome was measured on the BDI and HRSD. Inter-rater reliability for the PACS-SE subscales was reported to be quite variable (.38 - .77), and this may reflect the marked difference in experience reported for the two judges. This study distinguished the predictive power of adherence and competence ratings and also supportive and expressive competence, and so goes some way to addressing problems with the global approach identified in the IPT studies.

The frequency of techniques, as measured by the adherence subscales, did not predict symptom outcome, and this is similar to Rounsaville et al's (1981) finding on the proportion of time given to specific interventions, and Ogrodniczuk & Piper's (1999) finding that adherence on interpretive and supportive techniques measured with the ISIT were not related to outcome. Competent use of techniques did however predict subsequent symptomatic gains, but only for expressive techniques and not for supportive interventions. Supportive competence did however demonstrate a reasonable non significant effect size (.21). This finding was maintained when initial symptom gains, initial HRDS severity and expressive adherence were controlled (-.57, $p=.005$). It was of note that the two supportive scales had demonstrated the poorest interrater reliability, but when the scores were recalculated to include only those with adequate correlation the results of the primary analysis were unchanged. This result was also maintained when therapeutic alliance, measured on the Helping Alliance self report questionnaire (HAq; Luborsky et al, 1985), was added to multiple regression analysis prior to the technique variables.

These results appear robust in the face of a number of hypothesised explanations, but it should be noted that this demanded repeated multiple regression analyses on a small sample (twenty nine cases and four therapists), which increases the chances of type one error. This is a feature of many of the reported studies in which individual session ratings are used for relatively small patient and therapist samples. It is of interest however that unlike O'Malley et al (1988) competence was related to a standardised self report measure of symptom outcome, the BDI. It would have been of interest to know whether the reported finding were upheld if the independently rated HRSD had been employed as the outcome measure. Given the self report nature of the BDI it can be vulnerable to being used to communicate indirectly on general satisfaction with treatment or alliance, although such a bias would be unlikely to emerge on only one sub scale. As with previous studies the therapists involved were being trained, which may effect the variance in ratings. It would therefore be important to replicate these findings with a larger sample of experienced therapists to verify the conclusions drawn.

Svartberg & Stiles (1992, 1994) investigated the relation between adherence, competence and outcome in Short Term Anxiety Provoking Psychotherapy (STAPP; Sifneos, 1979) with patients with anxiety based disorders and produced markedly different findings in two studies. The first study compared and combined patient-therapist complementarity and therapist competence in early sessions as predictors of symptom change to mid- and post-treatment. Complementarity was found to be a significant predictor of SCL-90 change but not of SAS-SR or on the Dysfunctional Attitudes Scale (DAS; Weissman & Beck, 1978) change scores. Therapists' competence, rated on the STCRF (Svartberg, 1989), was not found to make a

significant contribution individually or beyond that made by complementarity. This contrasts with Barber et al's (1996) finding that competence is influential beyond the impact of alliance, but may reflect the different measures employed and conceptualisation of interpersonal interactions within therapy. Barber et al (1996) used an 11 item self report scale and Svartberg & Stiles (1992) used two independent raters coding sections of therapy transcript with the Structural Analysis of Social Behaviour (SASB; Benjamin, 1974), and therefore may have addressed different components of the therapy relationship. As noted earlier even subtle differences of definition and focus can alter and on occasions reverse related findings.

Features of the competence ratings in Svartberg & Stiles (1992) are also of note, as only one rater was used and the mean scores suggest relatively low competence levels across the training therapists. Test-retest scores suggest that the competence ratings remained stable, and the rater was one of those responsible for previous work establishing the STCRP's psychometric properties, but as previously noted these are not entirely robust, and stability of ratings could be true of biased as well as accurate evaluations. Details of pre-, mid-, and post treatment mean scores were not provided, but would be useful to interpret how much variance is being explained by a relatively limited range of competence ratings. This may reflect the opposite of the IPT studies in which therapists were reported to be consistently high in their competence ratings. In addition the patient sample appears to have been highly selected with seventy nine referrals generating only thirteen study participants, calling into question the generalizability of the findings.

In a subsequent report on the same sample Svartberg & Stiles (1994) examined the individual and interactive predictive power of alliance and competence for outcome. Alliance was measured with the Facilitative Alliance Inventory (FAI), an eleven item self report measure. As in the previous report competence scores were weighted by client difficulty ratings, which were derived from items in the semi-structured intake evaluation interview conducted by the therapists. In this report both alliance and competence were shown to make a significant contribution to prediction of SCL-90 change, but not to DAS change. This is unlike the previous report in which competence did not hold a significant relationship to outcome. Of particular interest was the negative relation of competence to outcome, indicating that lower competence was beneficial in terms of positive SCL-90 change. This is in keeping with Piper et al's (1991) report of an inverse relationship with transference interpretations and improvement. The authors draw attention to the limited range of competence scores suggesting that this result is potentially illustrative only of the influence of poor to moderate practice. They also speculate that competence in this case may reflect more rigid adherence to model, but this would suggest a failing of the rating measure which purports to measure both adherence and competence, defined in terms of quality of intervention or the rating procedure. This scale produces a measure of technical competence, which combines identification of specific techniques and evaluation of their quality in the one score, and so may be vulnerable to previous comment on overly comprehensive single ratings. Carroll et al (2000) have argued against such global measures of competence, and given the reasonably consistent finding that adherence is a poorer predictor of outcome than competence (DeRubeis et al, 1989, Elkin, 1988), this may highlight a vulnerability

in the current measure. This in combination with the admittedly “modestly tested” (p30) alliance measure and small sample size threatens the validity of the conclusions.

4.3 COGNITIVE THERAPY

Cognitive therapy is another psychotherapy model which has produced conflicting results on the predictive validity of adherence and competence measures. Luborsky et al (1985) reported that the “purity” of therapy techniques employed by cognitive, supportive-expressive and drug therapists significantly correlated with outcomes across and within therapists’ caseloads. This was essentially an adherence-outcome analysis, with purity measuring use of intended techniques and avoidance of proscribed interventions. Segments of therapy tapes were rated for the extent to which they reflected core concepts in the therapy manuals. Patient and alliance variables were also examined but the authors concluded that therapy “purity” was still independently and significantly related to outcome, which was defined symptomatically and in terms of employment, legal and psychological status. It is interesting that this study of three distinct forms of therapy reaches a conclusion consistent with Frank et al’s (1991) finding for IPT, and may suggest the value of the purity concept and the more functional definition of outcome.

A subsequent study which focused exclusively on cognitive therapists, was Shaw et al’s (1999) evaluation of practice in the NIMH TDCRP. This study involved review of a much more substantial body of data than previously discussed investigations, typically rating nine of twenty treatment sessions and producing a sample of 302 sessions

across 36 patients. CBT competence was rated on the Cognitive Therapy Scale (CTS; Dobson et al, 1985) and adherence was rated on the CSPRS (Hollon et al, 1988), and all evaluations were made blind to final outcome scores. Outcome measures included the SCL-90, BDI and HRSD. Multiple regression analysis revealed that the CTS total score accounted for a significant and unique 15% of HRSD end scores' variance, and this was statistically distinct from the influence of pre treatment HRSD severity and ratings for adherence and facilitative conditions. In contrast the CTS made a non significant contribution to the explained end point variance on the BDI and SCL-90, once again illustrating the power of a change of variable definition in this equation. This is of interest as this is one of the few studies to report a significant relationship between competence and an independently rated measure of symptom change. Previous positive findings have related to self report and functional measures of outcome, with the exception of the apathy score on the HRSD for IPT (O'Malley et al, 1988).

To further illuminate the relationship between competence and outcome the competence score was split between CBT skills and structure. The analysis identified the structure of the CBT as the more influential factor, and this was again demonstrated to be independent of adherence and facilitative conditions. When outcome was redefined as the presence or absence of clinically significant improvement on the HRDS and BDI the significant effects for the CTS total score disappeared but were retained for the structure subscale. This suggests that it was the competent structuring of the treatment specifically, which distinguished patients with and without clinical improvement. This is consistent with DeRubeis & Feeley's (1990) finding that one set of "concrete" CBT techniques, which correspond to the structural components of the

treatment, predicted outcome when measured early in treatment. Once again partial support for a component of competence and a specifically defined outcome variable is presented in Shaw et al (1999), and once again the functional definition of outcome i.e. clinically significant gain, is more fruitful.

It may be that the variability across and within therapists during the NIMH TDCRP aided the illumination of the competence – outcome relationship. The fact that CBT compared unfavourably to IPT in some analyses has been speculated to relate to the competence with which the CBT was delivered. There does appear to be some evidence that the therapies were delivered to different levels of competence, with 33% of the CBT tapes monitored precipitating red-line supervision calls, having been evaluated as falling below an acceptable standard, compared with only 3% of IPT tapes (Elkin, 1999). In addition when global ratings were made of how close therapists came to expert raters' concept of the ideal therapist in their respective treatments, based on video review and study participation data, IPT therapists came significantly closer to an ideal IPT standard than CBT therapists did to their standard (Elkin, 1999).

Milne et al (1999) also examined the effectiveness of CBT training with respect to client outcome by rating changes in therapists' competence and the corresponding changes in patient coping strategies after a forty day training program. Trainee therapists were rated at three points during a twelve session intervention, using the revised version of the Cognitive Rating Scale (CTS-R, Blackburn et al, 1999). The passage of time revealed a significant increase in overall therapist competence as rated by experts blind to the stage of therapy, and this reflected gains on eight of the thirteen subscales rated.

In line with this patients were reported to evidence a significant improvement in their coping skills as measured by the Coping Response Inventory (CRI, Moos, 1990). As presented, this study provides only correlational data, indicating an association between the two lines of change but no evidence of causal relationship. In particular it is of note that this evidence arises from a quasi experimental design which did not include a control group and consequently the results cannot safely be assumed to be specific to increased competence nor that greater competence does not reflect the facilitating effect of improving patients.

Overall then it appears that across a number of models of psychotherapy only partial evidence has been provided in support of greater competence of delivery predicting better outcome for clients. Subtle changes in definition across the relevant variables appear to exert a powerful influence over the relationships demonstrated. Patient rated change and functionally defined measures of outcome appear more likely to reveal a relationship with competence, although the former may also reflect dimensions of the therapeutic alliance and patient satisfaction with treatment. Global measures of competence, particularly those which collapse adherence and competence ratings, appear quite insensitive in the reviewed studies, and the limited range of competence which is often reported appears to further obscure the nature of the relationship.

5. THERAPIST FACTORS INFLUENCING PRACTICE AND OUTCOME

Luborsky et al (1986) suggested that "future research on the treatment performed by highly successful psychotherapists might tell us more about the process of change than the usual between group comparisons" (p511). This conclusion was based on his examination of individual therapist's success rates across four major outcome studies in which he found, "the frequency and size of therapists' effects generally overshadowed any differences between different forms of treatment "(p. 509). Specific therapist characteristics were not identified or hypothesised to be related to outcome, but rather the individual therapist's identities were included as random factors in covariance analyses on a number of outcome variables. Each reanalysis of the data revealed at least two significant univariate therapist effects, suggesting an unspecified trend for the therapist to be an influential factor in outcome. This finding is consistent with Shapiro et al's (1989) report on differential effectiveness in the Sheffield Psychotherapy Project, in which one therapist was found to be responsible for the apparent superiority of Prescriptive therapy over Exploratory therapy. Such results would suggest that the standardisation of practice in such studies has been unsuccessful in removing the influence of individual contributors.

These findings illustrate the importance of considering therapist factors in evaluating research findings and clinical outcome but do not clearly illuminate what is was about these successful therapists which produced these differential effects. In contrast other research has produced sobering findings on the potential for negative therapist effects e.g. Ricks (1974) studied the adult status of a group of adolescent boys treated by two

therapists and found that 27% of the “supershrink’s” patients were subsequently diagnosed with schizophrenia compared with 84% of the “pseudoshink’s”, having been comparable at the start of treatment. Obviously such a limited sample should be interpreted with caution and with consideration of other influential factors, but the point remains that practice and outcome differs across and within therapists. Therapist variables are typically collected less systematically than patient variables, and are rarely explicitly examined in routine practice, and consequently have been afforded less direct attention than other potentially influential factors. A number of possible explanations of such effects have been considered, including the experience or training of therapists and their personal characteristics and therapeutic preferences. However it must be borne in mind that the effect of such factors may not be stable, perhaps because of their interactions e.g. Orlinsky (1986) noted that therapists who produced poor or average results with some patients produced considerable improvement in others. Schaffer (1982) separated therapists’ influence along three dimensions, which included the therapeutic technique employed, skillfulness and the therapists’ personal qualities and interpersonal manner. The first two dimensions have already been addressed in previous chapters, and the third will be the focus of this discussion.

5.1 THERAPIST CHARACTERISTICS

Luborsky et al (1985) examined a number of factors thought to influence therapists’ success, among these characteristics of the therapists’ themselves. Based on prior work and reviewed expert opinion they developed a rating scale to evaluate three qualities which had been demonstrated to be predictive of outcome – therapists’

adjustment, therapists' skill and therapists' interest in helping patients. Therapists were rated by three independent, clinically experienced judges who demonstrated high interrater reliability (.89 or above on all items). It is unclear exactly what the ratings were based on, being described as familiarity with the therapists' work, which leaves some room for speculation on the focus and validity of the ratings. Factor analysis revealed two components. The first was characterised as "interest in helping patients" and the second smaller factor was "psychological health and skill", which appears to collapse two apparently independent concepts. Both factors revealed moderate non significant correlations with seven month outcomes variables, defined in terms of drug use, legal and employment status and psychological function.

Failure to achieve significance was attributed to the small sample of nine therapists by the authors, but vulnerability in the rating procedure must also be considered. The eleven items on the rating scale appear to be wide reaching and relatively ill defined in their descriptions, which may make them insensitive to the dimensions under study. In addition they would demand a significant and diverse knowledge of the therapists' work to make valid responses e.g. "unusually interested in helping patients", "unusually good adjustment", and "very capable and skillful therapist". Although the high interrater reliability would suggest that it was possible for the judges to make consistent ratings, such evaluation by peers may be influenced indirectly by an impression of the therapists' success i.e. those who were perceived as being more successful were rated as more skilled, interested and well adjusted. Raters were described only as being independent of the study but would by definition have to have been familiar with the therapists in order to make their ratings, making the findings potentially circular. This

would not explain the modest correlations however, as such a circular process may be expected to demonstrate a stronger relationship.

Blatt et al (1996) took Luborsky's ideas forward by examining the characteristics of effective therapists with the larger sample who participated in the NIMH TDCRP. Rather than trying to relate specific qualities to dimensions of outcome, therapists were rated for effectiveness and compared with more and less effective peers. A composite score of therapeutic change was derived from the five main outcome measures employed – HRSD, BDI, GAS, SCL-90 and SAS, and three groups of therapists were identified by splitting the sample into thirds based on this score. Those therapists who were identified as most effective were demonstrated to yield significantly more therapeutic change than the less effective therapists, while the moderately effective therapists were not significantly different from either their more or less effective peers. More effective therapists were also shown to have significantly less outcome variance than the other two groups, supporting the validity of the potentially arbitrary three way split.

A range of demographic and professional details was employed to characterise the therapists, along with a self report of attitudes and expectations with regard to etiology and treatment of depression. Demographic and experience variables did not differentiate the three groups of therapists, but profession and treatment orientation did. A significantly higher percentage of clinical psychologists than MDs were in the more effective group, and effective therapists typically treated depression exclusively with psychotherapy more often than with medication alone or in combination. It was of

interest that all three active treatment conditions – IPT, CBT and Imipramine plus clinical management, were represented in the more effective group, suggesting that the power of the therapist translates across treatment regimes as Luborsky et al (1986) suggested. It was also of interest that very few of the therapists recorded attitudes and expectations were significantly related to outcome, other than the lack of emphasis on medication in the more effective group. This may suggest some discontinuity between expressed attitude and practice, as it would be expected that beliefs about etiology and essential components of successful treatment would hold a stronger relationship with typical practice and therefore with therapeutic change. In summary this study appears to offer reliable support for the impact of individual therapists, and particularly their treatment orientation, in a well conducted research trial, but challenges connections which have been reported between outcome and years of general clinical experience.

In an earlier study Lafferty et al (1989) had also demonstrated that both in therapy variables and therapists' values could be used to identify more or less effective therapists. Trainee therapists were split into two groups (more and less effective), defined in terms of patient' response on the SCL-90, and assessed on an array of standardised measures. Therapist empathy, and therapists' evaluations of patient involvement in therapy and their own directiveness, successfully discriminated 78.57% of therapists. In a secondary analysis self reported values relating to the comfort and excitement in life and the importance of being intellectual successfully discriminated 83.33% of therapists. This is potentially consistent with Luborsky et al's (1985) finding on "interest in helping patients", as the less effective therapists were characterised by

Lafferty et al (1989) as being more self interested, reporting greater interest in their own prosperity and stimulation, which was hypothesised to interfere with their empathic and supportive capacity as values were ranked in order of importance. Unlike Luborsky et al (1985) therapists' psychological adjustment was not found to be significantly related to outcome, which may reflect differences in rating methods as in the earlier study this was rated by peers while Lafferty et al (1989) employed a self report format. Lafferty et al's (1989) study was relatively limited using only trainee therapists each evaluated with reference to outcome for only two patients on a single outcome measure. This was particularly important given that eleven therapist variables were included in the primary analysis and thirty-six value ratings were employed in the secondary analysis, potentially increasing the chances of type one errors. This study also reported only on the capacity of certain variables to discriminate between the two groups of therapists, but did not explore the relative contribution each made in explaining the outcome variance.

5.2 THERAPIST EXPERIENCE

The relationship between therapists' experience, regardless of therapy model, and psychotherapy outcome has received considerable attention and has produced conflicting results (Smith & Glass, 1977, Strupp & Hadley, 1979, Balastrieri et al, 1988, Stein & Lambert, 1984). Variations in how experience is defined e.g. by professional qualification, number of patients treated, years of practice, experience with the client group or with the therapy model, may significantly influence results. It is also likely that the relationship between therapist experience and outcome is not linear (Auerbach & Johnson 1977), being influenced by patient and relationship factors. Patients with more

complicated or chaotic difficulties might be expected to gain from the ability of an experienced therapist to recognise and adapt to these challenges while continuing to engage the patient in a way that less experienced therapists may struggle to do.

Early meta analyses which concluded that experience and outcome held no relationship (Smith & Glass, 1977; Shapiro & Shapiro, 1982) are vulnerable to criticism because of the limited range of experience reported. In the latter review therapists were reported to have a mean of less than three years training and experience, which can hardly be considered a robust test of clinical expertise. This is illustrated in Orlinsky & Howard (1980) who found that it was therapists with more than six years experience who demonstrated greater efficacy. Similarly Crits-Christoph et al (1991) classified therapists with extensive experience as those with more than five years post-doctoral or post-residency experience, and in their meta analysis experience was found to be a significant independent predictor of average therapist effect. It was of note that almost sixty percent of their sample had more than five years experience. Burlingame et al (1989) also reported that clients in time limited therapy who had experienced therapists demonstrated superior outcome ratings when compared to those of less experienced therapists, and increased levels of relevant training were associated with lower attrition, recidivism and clinically significant change. Methodological weaknesses blight many of the reviews conducted e.g. Nietzel & Fisher (1981) rejected thirty seven of the forty one studies used in Durlak's (1979) review on methodological grounds. However this in itself is not an explanation of the failure to find a positive relationship as Berman & Norton (1985) excluded studies with inadequate methodology and still could not find evidence of a significant effect for therapist experience.

Stein & Lambert (1984) attempted to address the difficulties of different definitions of experience by coding the variable along three lines. One produced a summary score of training and experience, the second calculated the difference in years of experience between more and less experienced groups and the third employed a five point rating systems which categorised level of experience. Although the overall effect size remained zero, the individual calculations demonstrated some differential capacity to pick up significant correlations e.g. the second system was a significant predictor of more positive outcomes for more experienced therapists. This highlights the importance of definition in interpreting results. These positive results were replicated in Stein & Lambert's (1995) meta-analysis of trials of clinical problems, which reveal modest effect sizes for experience on symptoms and client rated outcomes.

Overall it seems that despite considerable attention over many years it remains very difficult to draw any firm or general conclusions about the relationship between experience and outcome. Experience as a concept has been poorly and variably defined, making it both difficult to be clear exactly what is being measured and how the result may inform thinking in other models of practice and with novel groups. While intuitively appealing to think that experience generates more than a longer professional history it appears that this must be carefully defined before even modest results are revealed. Experience may be concluded to be too global a concept and it may be more fruitful to redefine the concept in terms of models of professional development e.g. Hogan (1964) and treatment orientation and skill.

5.3 THERAPEUTIC PREFERENCE

Robinson et al (1990) conducted a comprehensive review of the literature examining the evidence for psychotherapy as treatment for depression. One failing in the existing literature was noted to be a lack of attention to the therapeutic preference of the researcher conducting comparative investigations. Robinson et al (1990) noted that this had been shown to be an influential factor in previous literature, and independent judges were enlisted to rate the investigator allegiance for each of the comparative studies examined. Such allegiance was not always explicitly stated and consequently some ratings relied on interpretation of subtle indicators of preference. Impressive inter-rater reliability was reported using this method (.95), although this cannot be taken as a measure of inference validity.

This review uncovered considerable evidence of theoretical preference for specific forms of therapy among the researchers reviewed. This is perhaps not surprising given that it is often the very investment that researchers have in a theoretical tradition or its clinical application which, at least in part, drives them on to investigate questions of efficacy and mechanisms of change etc. Importantly Robinson et al (1990) demonstrated that this theoretical preference was significantly related to the reported outcome, such that the practice of preferred therapies tended to predict superior results to the practice of alternative therapy modes. This was found to be true, both for the allegiance reported within the study and when previous statements of allegiance were considered as a means of substantiating the allegiance rating. Robinson et al (1990) examined the relative impact of this association by means of a regression analysis, which partialled out the investigators' preference but retained therapy model. They

found that the treatment effect sizes were substantially reduced, leading Robinson et al (1990) to conclude that, on the basis of this analysis, there was no longer any evidence to support the relative superiority of one treatment form over another, and allegiance provided a superior explanation of outcome variance.

Theoretical preference is in itself a complex concept, potentially reflecting the credibility with which a treatment approach is viewed, the extent to which the rationale is explained and understood, and the clinical experience and familiarity the therapist has with a particular mode of treatment. Such allegiance may generate the practice of better therapy or it may in itself be the result of positive results which have been published on outcomes for specific models. Allegiance may also exist along a continuum rather than as a single choice of preference i.e. there may be an acceptable range within which a selection of different treatments may be judged to be preferable to others. This may be even more likely in the routine clinical setting where therapists may be less invested in producing research findings on specific models of interest and more interested in having a range of suitable strategies available to support their work. To the extent that manuals may facilitate learning about and research on a model of therapy, and support practice with guidance and illustration, they may encourage the preferential adoption of a therapy mode, and consequently improve outcome.

Given the apparent power of this factor to change the evidence base for different treatments it was important to examine the way in which therapist preference influenced practice. Robinson et al (1990) did this by looking into the possibility that such researcher allegiance may influence the quality of the therapy being conducted, a

factor which has been shown to correlate with outcome (Frank et al, 1991). This possibility was also noted by Hollon & Beck (1994), in their review of studies comparing CBT and behavioural treatments:

“....these studies were conducted by groups better known for their expertise with behaviour therapy; the cognitive interventions....tended to be simplistic in nature and the therapists, usually graduate students, relatively inexperienced with the approach” (p440).

Not surprisingly behavioural treatments were shown to be more effective under these circumstances, but it is also of note that the explanation collapses two potentially relevant concepts i.e. allegiance and experience.

Robinson et al (1990) examined the extent to which the therapies under study had been monitored e.g. with use of video or audio recordings, or had been conducted in accordance with therapy manuals and found that neither strategy resulted in reliable differences in absolute effect size between the studies. Unfortunately no information was presented on how the monitoring and manual instructions were implemented, nor the evaluation made of the therapists' performances to help interpret these results. Having a manual does not necessarily ensure that the directions are followed, nor does it automatically establish the quality of the therapy delivered, on model or not (Waltz et al, 1993). Without such information it is difficult to evaluate the significance of Robinson et al's (1990) conclusion that there is little evidence to support the use of treatments manuals as a means of increasing therapeutic efficacy or differentiating the

relative effectiveness of treatments. This would require more detailed information on the monitoring procedures implemented to ensure that the guidelines of the manuals had been followed and their efficacy. These findings do however support the proposal that therapists bring more to the treatment than the strategies outlined in manuals, and these inherent factors, however defined, have been repeatedly shown to influence the course and outcome of treatments.

Robinson et al's (1990) conclusions have however been challenged by Gaffan et al (1995) who examined the impact of therapist allegiance, on two samples of publications on CBT as a treatment for depression. This detailed investigation offered only partial support to the predictive power of researcher allegiance. Allegiance did account for a significant and additional proportion of the comparative effect size variance when treatment and control condition studies were analysed, but this relationship was substantially weakened when studies with two active treatments were reviewed. In these studies the effect of the treatment model remained, although it was reduced when allegiance was partialled out, unlike in Robinson et al (1990) where the treatment effect disappeared. Of additional note was the failure to find any relation between either comparative or pre-post effect sizes and allegiance, when a series of later CBT studies were analysed. The two samples of studies were not found to be different and it was suggested that allegiance might have a more powerful effect at an earlier stage in a therapy's published life. This would make intuitive sense as early publications would be more likely to emerge from highly invested groups, and consequently a potentially distinct group of therapists, than at a later stage when the therapy, its practitioners and target patients are likely to have diversified. It must also be noted

that Gaffan et al (1995) were specifically examining studies on CBT and Robinson et al (1990) were not similarly focused, and this may highlight the differential role of therapist allegiance in different models of practice.

It would appear then that the investment a researcher has made in a model of practice has some potential to influence treatment outcome. This however is important to understand in context. The studies examined explored the allegiance demonstrated by those researchers sufficiently invested in a specific model to complete empirical evaluations for publication. It is unclear whether the pattern of allegiance found in this sample would be representative of that found in unpublished clinicians who may work from a more eclectic base, or may prefer a number of models of practice. It is also unclear whether the allegiance rating attributed to the authors of outcome papers is an accurate reflection of the participating clinicians, as the samples cannot be assumed to be the same. The influence of the diversification of allegiance appears to be demonstrated in the second sample of studies reported by Gaffan et al (1995). It appears that as treatments become more established, the character of the practising clinicians changes, from that of the originating experts and allegiance become a less influential factor. It has also been proposed (Elkin, 1999) that there may be some overlap between experience and preference, with older more experienced therapists drawn to the established therapies, while younger therapists may be keen to work with recent developments, as was shown in the NIMH TDCRP therapist sample. If this were true generally it would be very important to attend to the potential confounding capacity of different therapist characteristics over others.

6. PATIENT PREDICTORS OF OUTCOME AND PRACTICE

Patient factors have not evaded attention in the literature produced by studies of standardised therapies. Interest lies in identifying those patients most likely to gain from interventions generally and specifically, and in offering guidelines for allocation of limited resources. Additionally such information may help to illustrate the mechanism of change operating within individual models by reflecting on the change that is effected. This is not a simple task given the diverse factors which have been used as a basis for patient selection e.g. motivation for insight (Malan, 1963), defence mechanisms (Buckley et al, (1984), social class and problem chronicity (Pilkonis al, 1984). This is made more complicated when the differential relationship between some patient predictors and some outcome measures after some therapies are considered (Elkin et al, 1985; Horowitz et al, 1984b).

When Piper et al (1985) examined a range of patient predictors of process and outcome in Short Term Individual Psychotherapy, defensive style and object choice of the patient emerged as the two best predictors of a number of patient and therapist rated measures of end point symptom severity and change. In contrast Sotsky et al (1991) examined a range of sociodemographic details, symptoms and diagnostic variables to predict independently and patient rated depressive symptoms at the end of treatment in the NIMH TDCRP. This study uncovered six patient predictors of outcome as well as treatment – predictor interactions. The best overall predictors were social dysfunction, cognitive dysfunction, duration of current episode of depression, double depression, endogenous depression, and expectation of improvement. Lower

or negative ratings on the first four predicted better outcome on the HRSD and BDI, while the same was true for higher or positive ratings on the last two. Lower social function, high initial depressive severity and impairment of function were shown to differentially predict superior response to IPT, while the same was true of lower cognitive dysfunction and CBT and Imipramine plus clinical management. While initially counterintuitive these results indicate the importance of providing a therapy, which can take advantage of patients' strengths, as well as helping them come to terms with their difficulties. Such consideration are not always readily apparent in routine referral practice in which patients with decimated social lives are referred for interpersonal therapy, and patients with overwhelming cognitive impairment are referred for cognitive therapy, potentially hampering both by imposing demands which the patient will be unable to meet.

These studies illustrate some of the difficulties in trying to draw conclusions from this evidence base. The use of different therapy models makes the generalisation of individual findings difficult, although the TDCRP addressed this by reporting overall and therapy specific effects. In line with different models are the corresponding conceptualisations of patient and outcome variables, which are clearly very different in the two reported studies, as with many others, and reflect the bias of the researchers. While this is not a fundamental criticism it does again impede the wider application of the reported results. In some studies e.g. Piper et al (1985), this can present a more serious problem. Two of the outcome measures employed were the patients' and therapists' ratings of the usefulness of therapy, which are vulnerable to being confounded by awareness of symptom change. Another common variation is the

diversity of diagnostic groups both across and within samples, often resulting in small, heterogeneous groups e.g. Piper et al (1985) reported on twenty seven subjects with ten distinct diagnoses. In order to find a route through the complexity of this picture this discussion will focus on individual factors in turn.

6.1 INITIAL SYMPTOMS

The interaction between illness severity and treatment response was substantiated in the NIMH TDCRP, when the sample was analysed with reference to their pretreatment symptom severity and functional impairment. It was in considering these patient variables as mediating factors to treatment outcome that IPT was shown to perform significantly better than the placebo treatment for the more severely depressed sample, a standard CBT did not meet (Elkin et al, 1995). The severity splits in the sample were based on two initial ratings – the HRSD and the GAS - providing a measure of depressive symptoms and general functioning, both from an independent rater's perspective. Analysis of covariance, which reflected initial severity and treatment group across four outcome variables, revealed considerably more significant results than the primary analysis for the whole sample had. The findings clarified that the less severely symptomatic or functionally impaired group responded equally well to all treatments, including the minimally resourced placebo and clinical management condition. The more severely symptomatic group not only required more active treatment, but also responded preferentially to medication and in a less consistent manner to Interpersonal Psychotherapy. In addition the HRSD initial score was revealed as a better predictor of IPT outcome than the GAS, while both initial ratings clarified the superior response of Imipramine over placebo in the more severe group. Such findings are of value in

determining resource allocation, and suggest that packages of formal psychotherapy, particularly IPT, may be best reserved for more symptomatically distressed patients. The quality of the CBT conducted in this study has been called into question and this may offer some indication as to why this group was not also significantly better than the placebo condition (Elkin, 1999). It should also be noted that CBT patients were not significantly inferior to IPT patients in final outcome scores, although this was the trend of the data. Given the superior response of patients with lower cognitive dysfunction to CBT it would be of interest to explore whether patients with reasonable cognitive function but more severe symptoms of another nature would perform as well as those who received IPT or Imipramine.

When Shapiro et al (1994) compared the effectiveness of CBT and Psychodynamic–Interpersonal Psychotherapy (Hobson, 1985) for subjects with different levels of pretreatment depressive severity on the BDI, they did not find significant differences between treatments on most outcome measures. The only exception was on the Present State Examination which revealed an advantage for CBT with moderately depressed patients, while PI demonstrated a marginal advantage for patients with more severe symptoms. These findings are consistent with the pattern of scores revealed in the NIMH TDCRP.

Piper et al (1985) employed more individualised measures of initial severity and outcome and found that initial severity of disturbance as rated by patients, therapists and independent raters did not consistently correlate with outcome. Patients were asked to describe their objectives for therapy and they, their therapists and

independent raters, rated their initial and end point problem severity, degree of change and made judgements on the usefulness of therapy on 7 point Likert scales. Unfortunately the reported results are not sufficiently detailed to illuminate the exact nature of the established relationships, indicating only that none of the initial ratings were significantly related to more than one of the outcome variables. This suggests some relationship but not a consistent one, and seems to offer less robust support to the importance of initial severity than Elkin et al (1995). This may reflect therapy differences, as CBT and IPT were also distinguished, although not significantly, in the TDCRP. Differences in conclusion also seem very likely to reflect the manner of collecting the defining information. As already mentioned it seems likely that ratings of the usefulness of therapy will be influenced by the extent to which target problems are estimated to have changed, making it difficult to tease out what is being measured in the end.

Stewart et al (1998) examined the pattern of symptoms in the TDCRP sample and found that atypical features helped to tease out treatment differences in outcome that had not been apparent in the primary analyses. The active medication was superior to placebo only for those patients without atypical features, and failure to take into account diagnostic subcategories led to an underestimate of the impact of this treatment for patients with typical depression. This patient variable did not predict differential outcome between the psychotherapies. This further highlights the predictive power of patient variables, which can be readily missed when the patient sample is evaluated only in terms of global diagnostic categories or retention in therapy e.g. completers or intention to treat, as is common place in statistical analyses.

Thase & Friedman (1999) also examined the impact of depressive subtypes on treatment outcome, and in particular the value of psychotherapy as a treatment for patients with melancholia and other severe depressive states, in their treatment review. They concluded that while there was some evidence of some melancholic and endogenously depressed patients responding to IPT or CBT this seemed to rely more strongly on a combined treatment approach with medication for good effect (Prusoff et al, 1980; Blackburn et al, 1981). Particular physical markers such as disturbed sleep neurobiology were poor prognostic factors for psychotherapy (Taylor et al, 1999).

When interpersonal problems have been taken as the categorical patient variable, the relation to outcome has not been consistent. The Inventory of Interpersonal Problems (IIP, Horowitz et al, 1988) is a commonly used measure of interpersonal functioning, which has been repeatedly related to treatment outcome. Studies using this measure have suggested that the character of the interpersonal problems experienced may predict response to treatment. Patients with "overly nurturant" problems gained more from psychotherapy (Horowitz et al, 1993), while ratings on the dominance-submission axis were not related to outcome for outpatients (Filak et al, 1986), but were for inpatients (Davies-Osterkamp et al (1996). Schauenburg et al (2000) similarly found that dominance ratings had no influence on outcome of symptom scores but affiliation ratings demonstrated a modest relation to outcome. Paivo & Bahr (1998) did not find that initial interpersonal problems were related to final outcome following experiential therapy but did find shifting patterns of association with dimensions of alliance across the course of treatment. As with previous findings on interpersonal problems it was not

global rating in this area which were important but the individual dimensions at specific times e.g. hostility and early alliance and social anxiety and middle therapy collaboration.

Sotsky et al (1991) reported better outcome for IPT patients with lower social dysfunction in the TDCRP, and this may again point to the importance of matching patient resources with therapy demands. Johnson et al (1994) examined the course of treatment response for patients who became depressed following a severe interpersonal event, those with non-interpersonal stressors and those without antecedent stressors. This did not reveal an overall difference in response to CBT, but those with any form of stressor were shown to have a faster onset of symptoms during the index episode, and those with interpersonal stressors specifically, demonstrated a significantly faster response to treatment than the other two groups. It was also of note however that those with interpersonal stressors reported a shorter mean duration of current episode than those without stressors, which may suggest that there is a benefit to early treatment for patients with acute onset symptoms in response to discrete interpersonal stressors.

Overall it appears that the nature and pattern of early symptoms is influential in the course of treatment, but this must be conceptualised both in terms of severity and symptom cluster. In addition the use of patient variables in one domain i.e. social functioning, to predict outcome in another i.e. symptom outcome, may be vulnerable to issues of pace of change and targets problems within different models of practice.

6.2 COLLABORATIVE CAPACITY

The interaction between patient difficulty and therapist performance was also addressed within the TDCRP, in order to understand more fully the interaction between factors which impede or enhance the administration of the treatment and the process assessment of therapist performance (Foley et al 1987). In this analysis of 35 IPT training cases, patient difficulty was defined not in terms of symptom severity but with reference to the patient's ability and willingness to establish a relationship and engage in the tasks of therapy. A difficult patient was characterised as hostile, defensive and help-rejecting. Repeated measures were taken of symptom severity, social adjustment and patient's attitudes and expectations of treatment outcome, as well as supervisors' ratings of patient's hostility on the Vanderbilt Psychotherapy Process Scale, and therapists' self ratings of therapy performance. Analysis of pre-treatment variables found no correlation between general and depressive symptom severity or social adjustment and patient difficulty, while this was apparent for patient's pre-treatment expectations of therapy, which were significantly correlated to patient difficulty. In addition both supervisors' and therapists' ratings of therapist performance were inversely correlated with ratings of patient difficulty. This process analysis is very informative in confirming that IPT therapists potentially work just as well with more symptomatic patients as less symptomatic, which is in line with Elkin et al's (1995) findings on symptom severity and outcome. An additional factor to be considered however is the mutual willingness and capacity of therapist and client to engage in the therapeutic task to facilitate the successful journey through therapy, and patient difficulty appears to impede this process.

Goldfried et al (1997) also found this continuity to be important in their content analysis of CBT and PI. They noted that high-impact sessions, which might be defined as a mini-outcome, were those in which the therapist worked within the prescribed framework and the patient was receptive. Low- impact sessions were characterised by the therapists' attempts to maintain the therapeutic focus being thwarted by the patients' resistance or non-compliance. It does seem possible however that these findings reflect a reciprocal client-therapist relationship, with patients not simply facilitating or thwarting adherent practice, but variance in the therapists' continuity also drawing different responses from the patient. Piper et al (1985) also found very similar results, when they examined patient predictors in terms of process in short term psychoanalytically oriented psychotherapy. The patient's defensive style, quality of relationships with important objects and psychological mindedness were found to be important predictors of favourable process, defined as patients providing important private material and the degree to which the patient understood and worked with the therapist. Of particular interest was Krupnick et al's (1994, 1996) findings that not only were therapeutic alliance ratings only significantly related to outcome for IPT patients in the NIMH TDCRP, but it was the patients contribution to the alliance and not the therapists which was significantly linked to treatment outcome.

These studies underline the importance of the engagement between therapist and patient in conducting therapy. The dyad undertake the work of therapy in the context of a relationship, not as unconnected individuals, and the interaction between their respective qualities, attitudes and skills are as important as their respective contributions in isolation.

6.3 PATIENT EXPECTATIONS

Foley et al (1987) examined the impact of patient held expectations of therapy on therapist's self-ratings and supervisors' rating of performance across 16 IPT sessions. Only therapist self ratings increased as pre treatment patient expectations increased i.e. became more positive, while supervisors' ratings were not related to this pre treatment patient variable. This differentiation may at least in part reflect the different focus of the two sets of in therapy ratings. Therapists provided a rating of the effectiveness of each session on a seven-point scale. Supervisors completed a much more detailed assessment of performance, which included a rating of the overall quality of the assessment but also a range of ratings on the strategies and techniques which were employed, producing a composite index of therapist' skill. Thus these two conceptually independent evaluations would not necessarily hold the same relationship to patient's pre-treatment attitudes. This may however indicate that supervisors are more able to separate skill from impact, and therapists may be influenced by the extent to which they regard their patients being satisfied or responsive to their interventions. It would appear that pre-treatment attitudes continue to exert an influence over the course of therapy such that anticipation of failure in the patient creates a corresponding sense of impotence in the therapist.

Hardy et al (1995), in another study of standardised psychotherapies, also studied patient's expectations of treatment. Depressed patients were asked to rate their expectation of psychotherapy prior to allocation to either eight or sixteen sessions of cognitive behavioural therapy (CBT) or psychodynamic-interpersonal therapy (PI), immediately prior to starting therapy and immediately after their first session. It was of

interest that patients rated the treatment principles of CBT as more credible than PI prior to allocation, but this balanced out when patients had a link with either form of therapy, through allocation or actual experience. In terms of outcome, it was endorsement of either treatment form that predicted PI outcome, while CBT outcome was not correlated with patients' ratings of treatment principle credibility. The authors suggest that the endorsement of treatment principles may indicate a psychological mindedness, which PI relies upon more than CBT, which in turn offers more direct teaching on the treatment rather than relying on what patients may bring to the therapy. It may be that models of treatment such as CBT, which actively educate the patient about the process of therapy, have greater scope for engaging patients who are initially uncertain about the value of treatment than less education based therapies. Consequently such models may be less vulnerable to the difficulties outlined in Foley et al's (1987) study of IPT training cases. IPT, while structured and focused, does rely upon a degree of emotional and interpersonal insight, consistent with its psychodynamic origins, which CBT may not require to the same extent.

Elkin et al (1999) tested the importance, in terms of engagement in therapy, of a match between the patient's treatment assignment and their beliefs about the cause of their problems and what will be helpful in therapy, using data from the NIMH TDCRP. Consistent with Hardy et al's (1995) finding, patients in the NIMH TDCRP were more likely to rate CBT preferentially than IPT, although details are not provided on whether this is pre or post treatment assignment. Unfortunately there were insufficient numbers to examine the impact of congruence for each treatment modality and only a psychotherapy and medication comparison was possible. They found modest support

for congruence between beliefs and assignment increasing the likelihood of patients remaining in therapy and establishing a positive therapeutic relationship, with nine of the eleven early terminators coming from the incongruent group. Outcome was not specifically addressed in this study but establishing a positive therapeutic relationship has been found to be positively correlated with better outcome (Hovarth & Symonds, 1991), as has retention in therapy (Elkin et al, 1989). Consistent with Foley et al's (1987) findings, measures of alliance were independent of initial symptom severity.

6.4 PERSONALITY FEATURES

Shea et al (1990) examined the effect of a comorbid personality disorder on the 75% of the TDCRP sample in which this was identified. The effect on outcome was not found to be uniform, such that mean depressive scores and work functioning at the end of treatment, were not significantly different for those with or without a personality disorder. Social functioning post treatment was however significantly worse for those with a personality diagnosis. These findings were demonstrated both for the presence or absence of any personality diagnosis and for individual diagnoses. Unfortunately analyses were not run to reflect the impact of number of Axis II diagnoses, as 57% of the sample were assessed to have more than one personality diagnosis. It was of interest that the two groups were not significantly different in rates of attrition, a factor that had previously been shown to relate to outcome (Elkin et al (1985) and therefore the results could not simply be attributed to the more diagnostically complex patients withdrawing from therapy earlier. The finding on social functioning is also intuitively acceptable, as patients with personality diagnoses might be expected to have had poorer social functioning prior to the depressive episode. If this were so, even with

good outcome this sample would be expected to fair less well than those who did not face such additional difficulties in this domain.

The question of diagnostic validity is important in understanding these results, and the data presented do not adequately support this assumption. Diagnosis of personality disorders was made on the Personality Assessment Form, which provides descriptive paragraphs to illustrate key features of each diagnosis as the basis for ratings on a six point scale. Diagnosis reflected a score of four or above. Diagnostic status was therefore based on single ratings, which is likely to be a less reliable measures of the underlying constructs than would be the case with more detailed diagnostic interviewing. Inter rater reliability scores are only provided for absence or presence of any diagnosis and for one cluster of diagnoses, as the reliability sample was unable to support further analyses. The scores do not suggest robust reliability in a notoriously difficult area, and rely on clinically inexperienced raters to make difficult choices on the basis of minimal guidance. It should also be noted that the validity of such diagnoses in the presence of a confirmed Axis I diagnosis must also be considered dubious, and could reflect the exacerbation of normally dormant personality traits which become more apparent under the strain of living through a major depressive episode. It is unclear then whether the results reflect personality disorders or underlying traits that were manifest in the context of acute depressive symptoms. Thus this study appears to offer uncertain evidence of the limited effect of personality disorders on the outcome in the treatment of depression.

Frank et al (1987) had also examined the influence of personality factors on response to IPT in a sample of patients with a recurrent depressive illness. The samples were not distinguished by outcome, as the design selected only those patients who had responded to combined treatment. Instead time to achieve remission was examined, and it was revealed that those who responded more slowly to treatment also demonstrated greater personality pathology on a wide selection of personality assessments. This does not offer a prediction of outcome but may be important in considering when to assess outcome, such that comparable symptomatic gains may be possible for both types of patient but will take longer for those with more pronounced personality based difficulties. It was also of note that remission was achieved for both groups by the end of a sixteen week program of IPT and medication. Some caution is necessary in drawing these conclusions however as personality assessments were made following symptomatic recovery. Retrospective analysis assumes that the greater prevalence of certain personality characteristics influenced the course of recovery rather than the length of treatment exacerbating certain key personality features such that they were more pronounced at the end of treatment.

Blatt et al (1995) and Barber & Muenz (1996) also examined the TDCRP sample to look at the power of specific personality characteristic to predict differential outcome for the two psychotherapies under study. Barber & Muenz (1996) reported that more avoidant patients gained more benefit from CBT than IPT, while the opposite was true for more obsessive patients. Factor analysis of the Dysfunctional Attitudes Scale revealed Perfectionism, was a poor prognostic factor on all outcome measures for all treatments, while the Need for Approval did not interact significantly with any of the

treatments (Blatt et al, 1995). While it is of interest to examine the predictive power of specific characteristics for specific treatments the reported results should still be interpreted with caution. In both studies the patients variables are identified with relatively minimal criteria i.e. ratings on single items or subscales from a single measure. The Barber & Muenz (1996) study is also vulnerable as only subjects from the completer sample, who also met criteria on the avoidance or obsessive items were included, which resulted in a small and potentially unrepresentative sample being used.

The above studies illustrate some of the difficulties reported by Mulder (2002) in his review of personality pathology and treatment outcome in major depression. The question of influence of personality pathology is not simple, and reflects multiple definitions, rated at different times in relations to different measures of outcome and treatment effectiveness. In addition the reciprocal interaction between personality characteristics and depressive symptoms i.e. being depressed for a long time may influence personality as much as personality predicts experience of depression, is offered inadequate attention in many studies. When flaws rather than design differences are controlled for it appears that only small effects are reported for this interaction (Mulder, 2002), and comorbidity between personality and depression should not be presumed to impede good treatment outcome, particularly in terms of symptom change.

It cannot be forgotten that patients presenting for treatment are more than the cluster of symptoms which trouble them. This is without doubt an important component to understand, as the evidence would suggest that different symptom patterns are both generally and in their individual interactions, differentially responsive to treatment. The

person who is troubled will also have a significant capacity to influence the course of treatment, and while this interaction may generate challenges which therapists may find difficult to manage at times, the presence of Axis II traits and diagnoses does not readily predict outcome. The reciprocal relationship which is established, and the mutual willingness to do this, provides the foundation on which any subsequent work will take place or will be rejected, and if this scene is not set the other factors are unlikely to have the opportunity to play out their parts.

7. MODELS OF SUPERVISION

Supervision may be conceptualised and evaluated in a number of ways. Function or goal, methods of delivery, correspondence with trainees needs, or effect on practice.

7.1 METHODS

Chevron & Rounsaville (1983) examined the impact of a number of supervision variables when evaluating the clinical skills of psychotherapists training in a new model of treatment. They conducted a comparative study of five methods of evaluation: 1. didactic examination, 2. global ratings by trainers, 3. supervisors' ratings, based on therapists' retrospective reports of therapy sessions in supervision, 4. therapists' self-ratings, and 5. independent evaluators' ratings of videotaped psychotherapy sessions. Each of these could be seen as means of testing therapist performance, but the results suggest that the method employed greatly influenced the results of the clinical evaluation. The authors found that the five ratings of clinical skills did not correlate in a uniform way, with significant agreement being found only between the individual supervisor's ratings and the trainer's overall ratings. It is of note that trainers also operated as supervisors, which may account at least in part for this agreement.

The didactic scores were not significantly correlated with other measures of adherence and in fact demonstrated a negative trend. This is a sharp warning against assuming that because a therapist can clearly identify appropriate interventions on an academic basis they will implement them in practice. It is of concern however that a negative association was identified, although this was not statistically significant, and may at

least in part be attributed to the small sample of therapist and cases. The authors conclude that this at least demonstrates that this method is inadequate to predict therapist's psychotherapy competence. This seems obvious given that it is essentially assessing academic ability in understanding and reporting concepts rather than interaction and therapy skills.

It is also of particular concern that the independent ratings and the supervisors' ratings were not correlated, and that supervisors were found to revise their ratings notably when subsequently shown video recordings of the therapy. It was of note that this discrepancy was seen to operate in both directions, with supervisors' ratings being seen to have both overestimated and underestimated the quality of the interventions based on the therapists' performance during the supervision meetings. This second rating was also found to correlate very well with independent raters. This finding is significant as it suggests that when faced with direct evidence of therapy performance, supervisors are as capable as independent assessors to evaluate competence and they do not continue to be unduly influenced by interpersonal factors relating to the supervisee or knowledge of symptom change. This may denote a change in focus in information gathering, with attention directed more towards the client in supervision and to the therapist in video review. This is meaningful in that it suggests that supervision would not necessarily have to be supplemented by an external evaluation to substantiate claims of adherence and competence, if supervision were based on direct observation of therapy. This study also highlights that just as the existence of a manual does not guarantee that its recommendations will be followed, neither does the existence of manual based assessment necessarily provide useful feedback to

enhance clinical practice. It appears that it is as important to base assessment on a close review of actual practice as it is to have specified criteria on which to make the assessment.

Typical supervision practice tends to rely on verbal review of therapy progress notes, and this is an important factor to consider when reviewing psychotherapy outcome research. Rounsaville et al (1986) argue that the advantage of using audio or video recordings, characteristic of practice in many studies evaluating standardised therapies, is that they provide an "accurate, permanent record of the entire therapy that facilitates the reliable evaluation of the content and quality of therapeutic interactions"(p365). This has to be weighed against the greater time involved in supervising in this way and some authors have examined the value of using segments of tape rather than the whole therapy hour in an attempt to combat this difficulty DeRubeis (1982). Such evaluations however only address adherence or capacity to distinguish models of practice and not competent delivery. It would seem then that the accuracy of assessing therapists' clinical practice does not just rely on the specification of technique, as provided by manual based assessments, but also the means by which this evaluation is carried out.

The function of this evaluation is also important to consider. If it is a straightforward evaluation of whether or not specific techniques were employed or avoided then an independent evaluation of therapy recordings may be adequate. If however the objective is a more dynamic process whereby the therapist is given feedback on their performance as a means of enhancing their adherence and competence then the

interactive sequence of supervision, again based on review of therapy recordings may be more appropriate.

In addition to considering the advantages, disadvantages and range of application of manual based psychotherapy the process by which it is integrated into a therapist's practice is of importance. Supervised casework is a common, if not ubiquitous, component of this process and like the standardised therapies it promotes, the strategies employed are varied. The models of conducting supervision are worthy of review.

7.2 MODELS OF SUPERVISION

There are numerous models of supervision, delivered by a variety of means. The divergence in strategy and conceptualisation may reflect the objectives of the supervision exercise e.g. to prepare therapists to perform a standardised model of therapy in a research trial or general training for a junior therapist or continuing professional development for therapists across the experience range. The level of clinical expertise already achieved by the therapists must also be taken into account e.g. a novice will require quite different support from an experienced clinician looking for specific input on a difficult case. The theoretical orientation of the supervisor and supervisee may also set boundaries on the exchange. The training demands of different models of therapy may also dictate that supervision is preceded by or run in conjunction with other learning processes. IPT, like other brief psychotherapies, requires therapists to have reviewed a manual of practice, and completed a period of didactic training prior to supervision. Psychoanalytic and psychodynamic training

requires that trainees concurrently attend educational seminars and undergo their own personal analysis while conducting supervised casework (Dewald, 1997; Binder & Strupp, 1997). The remit of this process is a matter of some debate, potentially relating to technical competence (Newman et al (1988), professional development (Guest & Beutler, 1988) and /or the ethical behaviour of students (Vasques 1992).

7.3 CBT AND DBT

Some forms of therapy employ a supervision model which parallels the therapy structure and process. Two examples of this are Dialectical Behaviour Therapy (DBT, Linehan, 1993a; 1993b) and Cognitive Therapy (CT, Beck, 1979). In DBT supervision, prospective therapists initially enter a presupervision phase during which they become oriented to the treatment e.g. through reading and discussion with practising clinicians, and enter into an agreement to commit to training, similar to the pretreatment phase of the therapy. The supervision then progresses through a series of stages, as with the treatment, starting with skill acquisition through course work, study and workshops, moving on to skills training in a group setting and then application with clients. The supervision is conducted in a group format, as with the therapy, and progress is assessed in an ongoing manner. The supervision groups operate within a number of agreed boundaries e.g. adhering to a dialectical philosophy, and with attention given to developing the skills which are promoted in the therapy setting e.g. emotional regulation, mindfulness and distress tolerance (Fruzzetti, Walt & Linehan, 1997).

Cognitive therapy supervision similarly maps cognitive therapy with a structured, focused and educational approach, for which therapist and supervisor are mutually

responsible (Liese & Beck, 1997). The supervision is typically conducted once a week for an hour and consists of an in depth review of a single case which may then be generalised to other examples. The authors suggest that when possible this should be supplemented with biweekly group supervision, which takes a more didactic approach and addresses common themes and difficulties. The structure of the individual sessions is similar to that of a therapy session with an agenda set, material being connected across sessions, review of homework and negotiation of new assignments, summary and feedback. The supervision is conducted with reference to taped therapy sessions which are reviewed in advance and standardised supervision instruments are employed e.g. the Cognitive Therapy Adherence and Competence Scale (CTACS, Liese, Barber & Beck, 1995). The content of the sessions also mirrors that of therapy with attention given to diagnosis of problems and application of appropriate cognitive models, and cognitive case conceptualisation. Within this model of supervision it is acknowledged that some teaching on basic counselling skill may be required. In this way the Cognitive model differs from that of IPT which assumes that therapists are experienced therapists and attention is more specifically addressed to shaping therapists' interventions to be consistent with the manual recommendations. In both examples interpersonal issues will be addressed e.g. therapist-client, therapist-supervisor, when they interfere with effective therapy.

In both reviewed examples it is possible to see that the supervision routine provides an, albeit limited, opportunity for the therapist to experience the different stages of the therapy itself. This resonates with the psychodynamic requirement that therapists undergo their own therapy, although as both CBT and DBT make very limited use of

transference interpretations therapists are not required to have addressed their own personal material during this process.

7.4 SUPERVISION FORMAT

The manner in which supervision is conducted is varied. The possibilities may reflect who participates i.e. an individual supervisee and supervisor or a peer group either alone or facilitated by a supervisor; the material used for review i.e. the trainee's process notes and session recall or an audio or video recording of the therapy session; when supervision is conducted i.e. following completion of the therapy session or while it is being conducted as in family therapy where live supervision allows the supervisor to interrupt the therapy with suggestions and comment; and how closely the process is linked with the therapy i.e. an independent supervisor who has no direct contact with the client or a co-therapist who operates as colleague and supervisor.

The relative advantages and disadvantages of these procedures have been debated in the literature, and the choice of format may reflect the factors already mentioned as well as the practical constraints on the participants. Group supervision may be more economical, financially and in time, but allows less time for each participant, and live supervision may only be available in centres with appropriate facilities i.e. two way mirrors and/or telephone contact but potentially allows trainees to progress to more complicated cases more quickly. Individual supervision based on process notes is the most commonly used format, based on ease and convenience, but may be vulnerable to incomplete and inaccurate reports on the part of trainees, arising from conscious or unconscious processes, and may be unduly biased by patient progress (Chevron &

Rounsaville, 1983). In some models of supervision however this process information is regarded as vital, rather than a disadvantage e.g. supervision in psychodynamic therapy reflects on the parallel processes which occur between therapy and supervision and it is the selection of material and way in which it is delivered during supervision which holds much of the valuable information. This would be less satisfactory, although not irrelevant, were the objective to rate the therapist's adherence to a novel model of therapy, when technical proficiency is more the focus. As Chevron & Rounsaville (1983) demonstrated ratings based on therapist process reports are weakly correlated with the ratings supervisors' make when recording of performance are available. In supervision with a clearly defined remit like this it is common practice for evaluation forms, such as those generated by the psychotherapy manuals, to be used to focus attention and for ratings to be based on video or audio recordings, at least for sections of the therapy. Under these circumstances the objective is generally not to teach the therapist the basic skills of conducting psychotherapy, as may be important in the more general practice of supervision, but to shape the use of interventions to be consistent with the recommendations of a particular approach. In more general supervision therapy specific forms would be used less often. The casework may span more than one model of therapy and more attention may be given to the therapeutic role and general skills e.g. ability to formulate difficulties within a psychological framework and fostering a therapeutic alliance, rather than model specific techniques.

Models of supervision may differentially attend to the context and wider organisational issues which impinge on the process. The prominence given to different strategies should reflect the function of the supervision exercise. This may be educative i.e. to

develop the skill, understanding and abilities of the supervisee; supportive i.e. to respond to the impact of clinical work on the therapists; or managerial i.e. a means of monitoring and controlling the quality of work conducted with clients (Hawkins & Shohet, 1987). Bernard (1997) extends this in a discrimination model of psychotherapy supervision in which she suggests three supervisory roles (teacher, counsellor and consultant) may be adopted in response to three different focus areas i.e. intervention skills, conceptualisation skills and personalisation skills.

7.5 DEVELOPMENTAL MODEL

The developmental approach to supervision (Hogan, 1964; Worthington, 1987; Stoltenberg & Delworth, 1987; 1997) which focuses on the changing supervision needs of the therapist over time has received considerable attention in recent years, and has been presented in several formats. This model is independent of any single psychotherapy model, tracing instead a generic supervisory relationship. Stoltenberg & McNeill (1997) present the integrated developmental model (IDM, Stoltenberg & Delworth, 1987) which proposes “using three overriding structures to monitor trainee development through three levels across various domains of clinical training and practice, thus integrating mechanistic and organismic models and providing markers to assess development across domains” (p 187) (Table 4). Specific recommendations are made for the supervision environment at each level in terms of degree of structure (decreasing with each level), appropriate client assignments, interventions to be employed and the mechanisms in operation.

Table 4: Integrated Developmental Model of Supervision

Developmental Levels			
Level	Motivation	Autonomy	Self-Other
			Awareness
1	High motivation High Anxiety Focus on skill acquisition	Dependent on supervisor Need for structure, positive feedback, minimal direct confrontation	Limited self awareness Focus on self: anxiety performance Objective self awareness Learns from outside source Difficulty seeing strengths and weaknesses
Transition Issue	May reduce for learning new approaches or techniques	May desire more autonomy than is warranted	Switches focus more toward client and away from self
2	Fluctuating; at times confident; More complexity shakes confidence; result is often despair , confusion and vacillation	Dependency Autonomy conflict; At times more assertive, develops own ideas Independent functioning; may want specific help Other times dependent and evasive	Can focus more on client; affectively empathic; understand worldview May become enmeshed and so not effective Issue is appropriate balance
Transition Issue	Increased desire to personalise orientation	Becomes more conditionally autonomous; better understanding of parameters of competence	More focus to include more reactions of self to client
3	Stable Remaining doubts not disabling Concerned with total professional identity and how therapist role fits in	Firm belief in own autonomy, not easily shaken; Sense of when it is necessary to seek consultation	Accepting of self, strengths and weaknesses High empathy, understanding Can focus on client process information, including own reactions Can use self in sessions

This model offers some guidance on the interaction between the emerging therapist and the process of supervision, and emphasises the dynamic nature of the exchange. The stages cannot be rigidly applied, as therapists may alternate between different

levels in both directions, and the respective developmental stages of the therapist and supervisor must be considered i.e. a novice therapist and experienced supervisor may operate quite differently from an experienced therapist and novice supervisor. The process of training and supervision may move the trainee from unconscious incompetence to conscious incompetence before completing the move to unconscious competence (Connor, 1999), with quite different therapist performance and related client outcome at each stage. It is therefore very important to consider timing when evaluating therapists and the supervision (Henry et al, 1993).

Such a model could be of great value in providing an integrative base from which to evaluate the developmental stage of the therapist in approaching casework. Therapists may be conceptualised to be at different stages of development in relation to different aspects of their work e.g. a level three therapist who initiates training in a new model of therapy may be more accurately described as being at level one in this component of their professional experience. This could provide a valuable foundation from which to review the confusing and conflicting findings which have been discussed on therapist experience, competence and outcome, and may provide a means of explaining the negative reports for skill and personal experience when therapists engage in therapy training (Henry et al, 1993). If the developmental model was combined with the specific adherence and competence measures previously discussed the two components could provide a more comprehensive evaluation of individual therapists, reflecting specific practice and more general training and support needs as suggested by Luborsky & DeRubeis (1984). While this may describe the ideal use of the available procedures, it should also be noted that this would typically extend far beyond

the remit envisaged for most individual measures of adherence and competence and may illuminate the parting of the way between research and general practice clinical supervision. It should also be noted that even if the developmental model were demonstrated to provide an accurate illustration of the progression of individual therapists it does not offer specific guidance on how to move through the transition issues, which would add to its utility greatly.

7.6 PROCESS MODEL

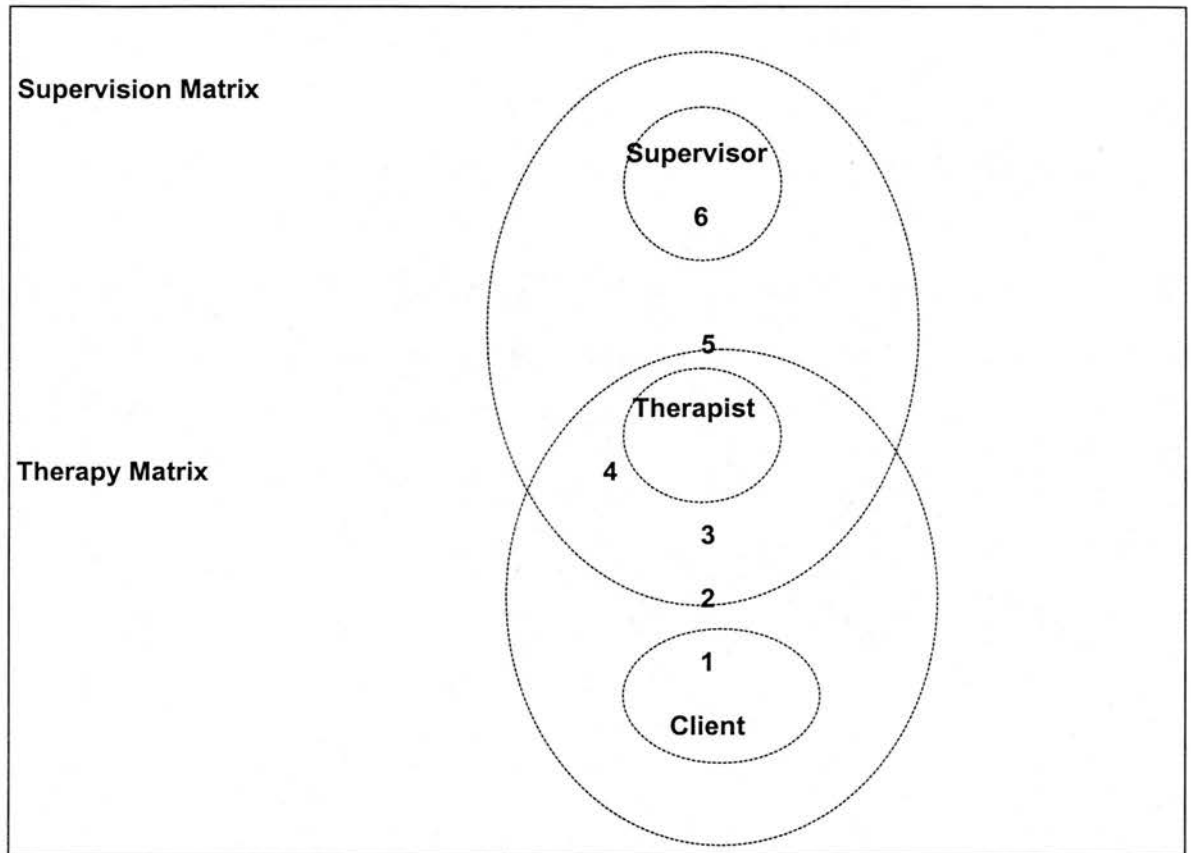
Hawkins & Shohet (1989) present a process model of supervision to represent the bias of attention within different forms of supervision, rather than developmental requirements of the therapist. They illustrate this with reference to two interconnecting matrices – the supervision matrix and therapy matrix (Figure 1).

They argue that a purely developmental perspective is vulnerable to being applied in an overly mechanistic way, relying on assumptions about the developmental route of individuals and their corresponding needs, without adequate attention to personal development within the individual, the style of the participating parties and the uniqueness of each supervisor-supervisee relationship.

Hawkins & Shohet (1989) suggest that supervision styles may be divided between two main categories:

- supervision which attends directly to the therapy matrix, such as that based on recall or recordings of the therapy session;
- supervision which attends to the therapy matrix as it is reflected in the supervision process.

Figure 1: Process Model of Supervision



The numbers in the diagram indicate the shifting focus of attention which may operate in different modes of supervision:

1. Focus may be directed to the content of the therapy session, with the supervisee and supervisor collaborating to reflect on the therapy exchange, either with reference to notes, recall or recordings.
2. Attention may be more specifically directed to the specific interventions selected by the therapist in terms of their nature, timing and rationale.

3. The dynamics of the therapy relationship may come under particular scrutiny, with review of the conscious and unconscious processes enacted during the exchange.
4. Supervision may attend more directly to the therapist's experience of the therapy in terms of personal material consciously or unconsciously stimulated, the role in which the therapist is being cast by the client, the therapist's attempt to counter this transference and the material projected onto the therapist by the client.
5. The supervisory relationship may become the focus of attention, as a mirror through which the therapy exchanges may be better understood.

The supervision and evaluation materials reviewed in previous chapters have focused more exclusively on the first two dimensions in therapy matrix. Hawkins & Shohet (1989) however argue that good supervision involves movement between each of the different modes. Awareness of the distinct modes may also facilitate a review of the supervisory exercise, both with the supervisor becoming aware of overused and underused strategies and by means of a common language through which both the therapist and supervisor can draw attention to specific areas. It is arguable that this process of feedback would be facilitated if it were enacted with reference to a standardised model of therapy with valid and reliable measure of the practice to inform the discussion. However it is also immediately apparent that the higher modes place increasing demands on the supervisor, which could be met by only by experienced and trained individuals. Such a comprehensive review of practice would move some

distance from the reliable but limited information supplied by some of the practice based evaluations, and may be in danger of assuming a greater breadth to the supervision exercise than may always be required or available. It may also be that the individual dimensions of the process model are not all met within a single supervisory relationship, but rather divided appropriately according to available resources, reflecting the reality of environmental and occupational factors. In such circumstances it may prove counterproductive if the replication of tasks confused or obstructed overall progress. It may also be however that by giving attention to the anxieties and self doubt which have been repeatedly demonstrated in the adherence and competence literature, and the way in which they become manifest in the therapy relationship, the supervision relationship at modes 3, 4 and 5 has scope for helping therapists to manage their personal response to learning and avoid objectifying their patients as more or less ideal candidates for their newly acquired skill.

It is clear even from this brief review that supervision is far from a simple concept, and the methods employed to monitor the application of standardised and manual based therapies offer only one potential model. As demonstrated in previous chapters discussing the therapy exchange, the supervision relationship is multiply defined and operates on many levels. This exchange can help us to learn about therapies, therapists and training procedures, and ultimately to gear resources to achieve good outcome by means of a constructive process for our clients.

7.7 AIMS OF THE STUDY

The study aims to examine the early casework conducted by therapists new to the Interpersonal Psychotherapy Model. In particular the study aims to review the application of formal adherence and competence measures in reviewing the completed casework. Such measures have been used in previous reports on IPT training and supervision but have been employed with samples of highly selected therapists and evaluators. This study will examine the use of these forms to evaluate the more diverse sample of therapists who are currently completing training in IPT.

Given that expectations of practice have been established with therapists who no longer reflect the majority of IPT trainees this study will explore the standards of practice demonstrated by therapists with different professional backgrounds and with a range of experience and therapeutic allegiances.

Patient variables have been demonstrated to exert an influence on treatment process and outcome and this study will explore the relationship between initial symptom severity, early response to treatment and problems in conducting therapy on therapy process and outcome.

This study also aims to explore the differences in practice observed between the most and least effective therapists and to relate this to competence ratings and strategy selection.

7.8 HYPOTHESES OF THE STUDY

1. Therapists will demonstrate adequate adherence and competence in conducting early IPT casework.
2. Competence ratings will show adequate reliability between the therapy supervisor and external raters.
3. Higher levels of competence will be related to the greatest reductions in patients' symptom severity scores.
4. Therapist performance and patient outcome will not be significantly influenced by therapists' reported therapeutic allegiance.
5. Therapist performance and treatment outcome will not be significantly influenced by therapists' clinical experience.
6. Patient generated problems in conducting therapy will be inversely related to competence and outcome ratings.
7. Patients with severe depressive symptoms at the start of treatment will respond to treatment as well as patients with moderate or mild depressive symptoms.
8. Patients who show greater early symptom change in treatment will report greater overall change than those who show less initial symptom change.

8. METHODOLOGY

The aim of this study was to examine the supervised casework conducted by a range of therapists following a series of introductory workshops on Interpersonal Psychotherapy. In particular the application of a series of evaluation forms, which specify criteria for adherent practice of the model and provide measures of the competent delivery of the specified techniques, were evaluated.

Data are presented from two main sources. The first series of therapists had completed didactic training in IPT at an international training centre and were undertaking casework in preparation for participation in a clinical trial comparing IPT and CBT in the primary care setting. This was a randomised controlled trial funded by the Scottish Office. The second data source was a selection of therapists who had attended one of six introductory courses in IPT run across the UK, and were preparing to use the model in their routine clinical practice. Patients were drawn from the therapists' local services.

The supervised work was conducted to shape and add to the existing repertoire of skills of therapists with a range of clinical experience, and to facilitate therapists in conducting IPT in a manner consistent with the manual guidelines.

8.1 PATIENTS

The study reviews casework conducted with twenty eight adult outpatients. Patients who were seen in the pilot stage of the primary care study were initially selected from the

existing clinical psychology primary care waiting list, to provide a sample of subjects representative of those proposed in the study.

Patients included in the main study were 18-65 year olds presenting with recent onset major depression, generalised anxiety disorder, panic disorder with or without agoraphobia, social phobia, mixed anxiety and depression, and dysthymic disorder. Patients excluded from the main study were those presenting with organic brain disorders, psychotic disorders, eating disorders, obsessive-compulsive disorders, personality disorders in the absence of a depression or anxiety disorder, patients with evidence of alcohol or drug abuse, and patients who had received ECT or in-patient care in the last 12 months. The sixteen patients selected for the pilot study cases all had a diagnosis of current major depression.

The remaining twelve patients who were being seen in different locations in the UK, were selected from the waiting lists serviced by the therapists. Four patients came from two clinical psychology primary care lists, two from a tertiary care waiting list, and six came from four multi-disciplinary mental health team lists. All patients had a diagnosis of major depression, and none met the exclusion criteria that operated for the primary care RCT sample.

A potential difference between those referred to the RCT therapists and those in routine practice arose on the basis of time on the respective waiting lists. The initial recruitment from an existing waiting list for the RCT pilot study was disappointing, with a disproportionately high DNA rate. This seems likely to have been a consequence of the

extremely long waiting list that the local clinical psychology service operated, and this resulted in a delay in starting the pilot phase for a number of therapists. Several therapists were forced to offer appointments for up to five patients before one attended. In order to overcome this difficulty, participating GP practices were approached directly and asked to refer to the study, with randomisation initially being made to the pilot study and then to the main study. This greatly improved the recruitment process and increased the number of participating patients.

In summary, in the pilot phase 73 patients were offered assessment appointments. This resulted in 28 subjects being recruited into the pilot study, from both the existing clinical psychology waiting list and direct referrals. 45 of the patients approached were excluded from the pilot study. These individuals were excluded for a number of reasons: 5 failed to meet inclusion criteria, 9 declined to participate having recovered since their referral was made, 5 wished to remain on the clinical psychology primary care waiting list rather than participate in a research study, and 26 failed to attend without prior notice. Details of sixteen patients seen as part of the IPT pilot are reported in this paper.

In contrast those patients being seen in the routine clinical settings were either prioritised over patients ahead of them on the waiting lists, as with the RCT sample, or offered appointments prior to being placed on the waiting lists. As a consequence of this early case identification all of these patients attended with their first invitation for therapy.

8.2 THERAPISTS

Seventeen therapists were involved in the study. All therapists undertook training in order to either meet the preliminary selection criteria for participation in the primary care study evaluating IPT and CBT in primary care or in order to provide IPT in their local service. The therapists represented a range of mental health professions and reported a wide range of clinical experience. Full details are presented in the results section.

8.3 SUPERVISOR

One supervisor acted as trainer for the majority of therapists, and supervisor to all. The supervisor was a qualified clinical psychologist, who had completed IPT training and supervision on ten individual cases prior to offering supervision. Additional casework had been conducted with between twenty and fifty IPT cases prior to the individual supervision exercises beginning. The supervisor had observed the work of several IPT supervisors, with a view to preparing to take on this role. The supervisor was a member of an international supervisors group, meeting once each month to review issues associated with supervision, and to standardise the practice of supervision. During the course of the study the supervisor was also made training and supervision co-ordinator of the UK IPT network and joined the training and supervision committee of the International Society of IPT (ISIPT). The supervisor had experience of teaching and training in IPT.

8.4 TRAINING AND SUPERVISION PROCEDURE

The training for all therapists took place in three distinct stages:

1. Review of the training manual

2. Didactic training
3. Supervised casework (Rounsaville, 1986).

8.5 TRAINING MANUAL

The training manual “describes the theoretical and empirical basis for Interpersonal Psychotherapy for depression and offers a guide to the planning and conduct of the therapy” (Klerman et al, 1984, p. ix). In offering this guide the therapist is assisted in discriminating between those techniques and strategies which are consistent with the prescribed approach, and those which may be part of their current practice but which are not used in the interpersonal model. The techniques to be used are operationalised and the relative importance of different strategies is explained for each of the three stages of therapy - initial, middle and termination. An outline, consisting of distinct goals and strategies, is provided for each of four focus areas, which provide a general structure for the work to be conducted in addressing the individual's interpersonal difficulties in the context of a depressive episode (Appendix 1). Distinct strategies are similarly provided for the initial assessment phase and the termination of therapy (Appendix 1). A range of appropriate techniques, not unique to IPT but consistent with its use, are described and guidance is offered on how to manage difficulties which may be anticipated to arise when using this model, e.g. patients seeing depression as incurable, involvement of significant others. The therapist and patient roles are also explained.

8.6 DIDACTIC SEMINAR

The taught courses were used as a means of explicating the details presented in the manual, and through case presentations, demonstrations and the use of role play exercises the therapists were encouraged to adapt their current practice to meet the demands of the operationalised description of the IPT model. This involved the modification of techniques already in their repertoire and the development of new approaches required by IPT.

Five of the therapists completed a four day didactic training course at an internationally recognised centre for training, the Clarke Institute of Psychiatry, Toronto. Ten of the remaining therapists completed one of six training courses conducted by the trainer/supervisor, who had been trained for this purpose at the original Canadian centre and had previous experience of training in IPT. Four of these courses were run in Edinburgh and were open to a wide selection of therapists. Three more were conducted in the local setting for the therapists (London, Cardiff and Dumfries) and were attended by a selected audience of local therapists. Both local training courses were conducted with a significantly smaller number of participants. The final therapist was trained independently at another introductory course in the UK.

8.7 SUPERVISED CASEWORK

Following completion of the first two stages each therapist was allocated cases for which he/she received regular written and face to face or telephone supervision. The aim was to provide a weekly review of the current casework, based on the review of audio recordings of each therapy session. This followed the suggestions of Chevron et al

(1983), who found a low correlation between ratings based on recordings of therapy and the more traditional practice of discussing case notes. All therapists were supervised by a single supervisor (who was also trainer to eleven of the therapists) while conducting therapy with depressed patients. There are a number of examples of IPT supervisors being used as practice evaluators in the literature (Foley et al, 1987; Rounsaville et al, 1983), and this reflects the embedded nature of IPT supervision and evaluation to date. This brings the advantage of a rich and comprehensive appreciation of the casework conducted but also the potential disadvantage of biasing effects from the supervision relationships and a generalisation of ratings across the treatment. In the present study the aim is to evaluate the suitability of using evaluation forms in the supervision context rather than the evaluation forms independently and therefore by design required regular review and dialogue.

Patients, who were diagnosed as having a current major depressive episode alone, were selected for the most part as it was for this disorder that the treatment intervention was originally designed. It was therefore considered appropriate that initial clinical practice should be with this group, rather than one of the groups for whom the model has been adapted.

8.8 MEASURES

The instruments used to assess the therapists' skill in implementing IPT strategies, as described in the IPT manual, were those employed on the NIMH TDCRP (Rounsaville, et al, 1986) (Appendix 2). These forms had been developed to provide explicit criteria for determining whether therapists had achieved and could maintain an adequate level of

adherence and competence in performing IPT. Eight evaluation forms were used in this supervision exercise, although the entire selection of forms was not used for every session, but rather reflected the stage of therapy being conducted.

The forms used were:

- Initial Sessions Checklist
- Therapists Strategy Rating Form (TSRF)
- Process Rating Form (PRF)
- Grief Rating Form
- Interpersonal Role Disputes Rating Form
- Interpersonal Role Transition Rating Form
- Interpersonal Deficits Rating Form
- Termination Sessions Checklist

The specific tasks and strategies of each stage of therapy were made explicit in the initial sessions, focus sessions and termination sessions forms. The TSRF reviewed the evaluation of the major problem area and selection of a consistent intervention focus, while the PRF detailed the range of techniques that may be appropriately applied across all of the sessions. This form also provided a rating of the frequency with which techniques inconsistent with an interpersonal approach were employed e.g. overly behavioural or psychoanalytic interventions. Consequently the initial sessions form was used for the first four sessions, along with the PRF. The middle sessions were reviewed using the TSRF, the focus form consistent with the formulation and the PRF. The last

four sessions were reviewed with the Termination sessions form, the TSRF, Focus form and the PRF.

Each specified task was rated firstly on whether or not it was used, and for those techniques not employed an additional rating was available on whether use of this technique would have been recommended. Not all techniques are intended for use in every session, and this recommendation allowed a distinction to be made between appropriate and inappropriate technique selection. This rating was used as a guide for supervision discussion and is not employed in the subsequent analyses. In addition to these adherence ratings a competence rating was made on the quality of the intervention, using a seven point Likert scale, ranging from outstanding (1) to poor (7). This took into account factors such as timing, appropriateness, and completeness of intervention. This is consistent with Waltz's (1993) recommendation that the quality of an intervention i.e. therapist competence, should be assessed in combination with rating the presence or absence of specific techniques. Purity of intervention was also assessed by rating the frequency with which techniques inconsistent with the manual instructions were employed (Hill et al, 1992).

Previous studies, which have used the IPT supervision forms, have reported adequate to good inter rater reliability scores. Foley et al (1987) reported intraclass correlation coefficients for the agreement between independent evaluators ranging from .60 to .80. This was based on the videotape review of each session of thirty-five training cases conducted by eleven therapists. Chevron & Rounsaville (1983) reported the overall inter-rater agreement on IPT competence scores for 27 sessions was $r = .88$ ($p < .001$).

This was based on videotape review of three sessions (1, 6 and 11) per case for nine therapists. Evaluations were made by the therapists' supervisors and an independent evaluator. Both evaluators had been involved in the therapists' training but then supervised a proportion of casework and independently evaluated the others. As previously noted these reports are based on more highly selected therapist and rater samples than is the case in the current sample. Rounsaville et al (1986) describe the independent evaluators as "being chosen from the group who had participated in developing IPT and compiling the training manual" (p 366).

As previously noted these forms are employed without an independent rating guide, relying on the IPT manual for definition and instruction. While this source is helpful it is not exhaustive and was not designed for this purpose. This is a key and central vulnerability of the existing IPT resources to guide supervision. One objective of the present study was to establish whether these rating forms could be used in a sufficiently reliable way by supervisors who were not as closely involved in their design as those employed in the TDCRP. Attention to supervision procedures focusing on adherence and competence has been largely neglected in the IPT literature since the TDCRP, and this has happened in conjunction with a rapid expansion in interest and use of the model across many countries. It is therefore very important to review whether the original guidelines are adequate to the current day task of facilitating and evaluating the casework conducted by a less selected sample of therapists when used by supervisors and evaluators with a range of personal training experiences and clinical interests.

In order to address this issue a sample of individual session recordings were selected for second ratings by one of two experienced and independent IPT supervisors. Both independent raters had been practising IPT for five years and had experience of conducting IPT training. The independent raters were blind to the identity of the therapists and had had no contact with the majority of therapists reviewed. It was possible that one therapist may have been recognisable to the independent raters although no information was provided. Neither independent rater had been involved in any of the therapists' training. Tapes were selected from early in the middle sessions in order to eliminate issues and tasks specific to the assessment phase of IPT and to minimise the impact of symptomatic gain which may bias ratings of adherence and competence. Tapes were used for sessions five or six for the majority of cases, and in three cases session seven was used, as an earlier and audible tape was not available. This resulted in the casework of thirteen of the seventeen therapists being second rated. Unfortunately the four other therapists had retained their tapes following supervision and had erased them for confidentiality purposes prior to the second ratings being made. The independent supervisors were blind to the selected focus prior to reviewing the tapes and to treatment outcome for the patients. They were informed of the session number in order to use this contextual information to facilitate judgements on the appropriateness of techniques and strategy selection for the stage of therapy.

All supervision forms were completed by the raters following audiotape review of the entire session. Sessions were rated weekly before the supervision session and ratings were blind to final outcome. While review of sequential sessions was likely to give the supervisor an informal indication of the progress patients were making this was

distinguished from a formal symptom review which was conducted off tape. In addition given that the evaluations were embedded in a supervision relationship, the continual review was necessary to allow appropriate and contextual corrective feedback to be given to therapists.

Recordings of the majority of sessions were reviewed for twenty three of the cases reported and at least six sessions for the remaining five cases. In the incomplete sets this was due to the tapes being inaudible or unavailable for review or therapy being terminated early. In these cases supervision was conducted with the available tapes and with discussion of the sessions for which recordings were not available. Cases in which there is an incomplete sample were dropped from some subsequent analyses. In order to make use of the ratings which were available for some sessions a performance analysis, as distinct from a competence analyses, is also conducted (Bein, 2000). This provides a greater number of cases for performance-outcome analyses.

Previous reports have indicated a high correlation between the TSRF and PRF ratings and have employed a composite score of overall competence (Foley et al, 1987). The index score is calculated as the mean of all TSRF, focus forms and PRF items related to the quality of the therapist's performance and is used as a measure of the overall and phase specific competence of application of IPT strategies, techniques and skills. A composite score is used in some analyses.

8.9 SYMPTOM CHANGE AND OUTCOME

The Beck Depression Inventory - II (BDI-II; Beck, 1996) was administered before a number of sessions to provide a measure of symptom change across the treatment and at the end of the sixteen sessions. All cases have baseline, mid-treatment (session six for the majority of cases) and end of treatment ratings. The BDI is a well recognised and commonly used measure of depressive symptomatology, with well established validity.

9. RESULTS

9.1 THERAPIST DETAILS

Details of clinical experience were collected on the therapists participating in the study (Appendix 4). In total six male and eleven female therapists, with an average age of 36.7 years (SD 6.9, range 24-46), took part in the training and supervision. Two therapists left their posts during the study and consequently some additional information is only available on a selection of the participating subjects.

9.2 CLINICAL EXPERIENCE

Of these subjects, four were qualified clinical psychologists, with at least a Masters level clinical qualification, and five more had undergraduate degrees in psychology and were employed as research psychologists, one of whom was also a qualified CBT therapist. Two of the clinical psychologists also held doctorate level degrees. In addition four therapists were psychiatrists, one was a clinical nurse therapist, two were community psychiatric nurses and one was a senior occupational therapist. Twelve therapists were employed in the NHS, and five were employed in Edinburgh University with clinical links in the NHS.

The mean duration of clinical practice was 8.3 years (SD 6.27, range 1-20). When experience with patients with depression was assessed, two therapists (11.8%) had seen less than ten patients with depression prior to participating, one (5.9%) had seen 10-20 depressed patients, and the remaining fourteen (82.4%) had seen over 20 patients with depression. Six therapists (37.5%) had completed formal training in another form of psychotherapy prior to participating in the study.

9.3 DIDACTIC TRAINING TO SUPERVISION

Only one of the therapists had attempted to use the IPT model prior to participating in this exercise. As the therapists had undertaken training at different sites and at different times, training experience was reviewed. Five therapists were trained at the Clarke Institute of Psychiatry in Toronto, eight were trained in Edinburgh, three were trained in their local area, and one was trained on another UK course. The mean delay between completing didactic training and starting supervised casework was 10.7 months (SD = 7.98, range 1-26). ANOVA revealed a significant difference between the therapists trained at different centres on delay to casework ($F(2,25)=7.7$, $p=.007$). Subsequent comparisons revealed that the Canadian trained group reported a significantly greater delay between completing didactic training and starting casework than the Edinburgh trained sample ($t=3.88$, $df=11$, $p=.003$), and approached significance in comparison to the other UK trained therapists ($t=2.18$, $df=6$, $p=.072$). This reflects the greater availability of training which emerged over the course of the study and the practical difficulties faced by the earliest trained therapists e.g. waiting for supervision to become available, repeated failure to attend in new patients, rather than a reluctance to complete the training and supervision exercise. Delay to supervised casework was not significantly correlated with baseline, end point or change in BDI-ii scores nor with maintaining an interpersonal focus in the assessment and formulation sessions, suggesting that the practical difficulties faced by some therapists were not reflected in biased client selection or associated with differential treatment process and outcome.

Therapists who had attended different training centres were also compared on client and therapist variables and this revealed two significant differences, firstly on baseline BDI-ii scores ($F(2,25)=5.27$, $p=.01$) and secondly on years of clinical experience ($F(2,25)=12.63$, $p=.01$). The Edinburgh trained group and the externally trained therapists differed significantly on baseline BDI-ii scores ($t=2.9$, $df=17$, $p=.01$), with the Edinburgh therapists working with more severely depressed patients ($M=35.9$, $SD=8.2$), than the externally UK trained therapists ($M=23.8$, $SD=6.7$). The Canadian trained therapists did not see patients who were significantly different on baseline BDI-ii from the other two groups ($M=29.1$, $SD=6.9$). The Canadian trained therapists however did report significantly longer clinical experience ($t=4.2$, $df=9.6$, $p=.01$) than the Edinburgh trained therapists, with the former reporting $M=13$ years ($SD=6.4$) and the latter $M=4$ years ($SD=2.5$). The external UK therapists did not differ significantly from either.

It should be noted that the externally trained group contained only four therapists, who saw five patients in total, and therefore 76% of therapists saw patients who could not be distinguished in terms of baseline severity. This difference may reflect a selection process which occurred when therapists were selecting among newly referred cases and presented for supervision when they had identified someone they believed to be a good IPT candidate, compared to the Edinburgh and Canadian trained groups who were more likely to work with the patients who were allocated to them and who turned up. Despite these baseline differences, therapists from different training centres did not produce significantly different treatment outcomes (F

(2,24) = 1.65, ns), when an ANCOVA was conducted controlling for baseline symptom severity, suggesting that training site was not independently related to outcome of treatment.

The difference in experience may also reflect the fact that the Canadian trained therapists were those responsible for initiating the use of IPT in Scotland in order to conduct a clinical trial, while the Edinburgh trained group were therapists who took up a local training opportunity when it became available. These two findings together do mean however that the therapists with the shortest duration of experience were typically seeing the most severely depressed patients in the sample. These differences will be taken into account in subsequent analyses.

Once supervised casework had begun therapists attended a mean of 17 (SD = 8.33, range 4-32) individual supervision sessions, extending over one to three overlapping cases. A mean of 13 (SD 5.6, range 7-16) therapy tapes were submitted for assessment per case.

9.4 PATIENT DETAILS

Twenty-three (82.1%) of the participating patients were female and five (17.9%) were male. Their mean age was 38.7 years (SD 10.3, range 22-56). Ten (35.7%) were single, eight (28.5%) were married or cohabiting, seven (25%) were divorced or separated, and two (7.1%) patients were widowed. Sixteen (59.2%) had gone on to further education at college or university, and the remaining eleven (40.8%) had a secondary school education. Nineteen (70.4%) were in paid employment at the time

of the study, one was a homemaker (3.7%), five (18.5%) were unemployed, and two (7.4%) were signed off from work as a consequence of their depression.

All of the patients met diagnostic criteria for a current major depressive disorder, and eighteen (66.7%) had long standing mood related problems, reflecting either a recurrent depressive illness or dysthymia. Sixteen (59%) of the patients had had at least one previous episode of major depression, at least three (11%) had a history of anxiety disorder and one (6%) reported a past diagnosis of bulimia nervosa. Eleven (40.7%) patients had previously been treated with anti depressant medication, and eight (29.6%) more had previous experience of psychotherapy, either with or without medication. The range of psychotherapy approaches reported included CBT, person centred counselling and psychodynamic work.

9.5 SYMPTOM CHANGE

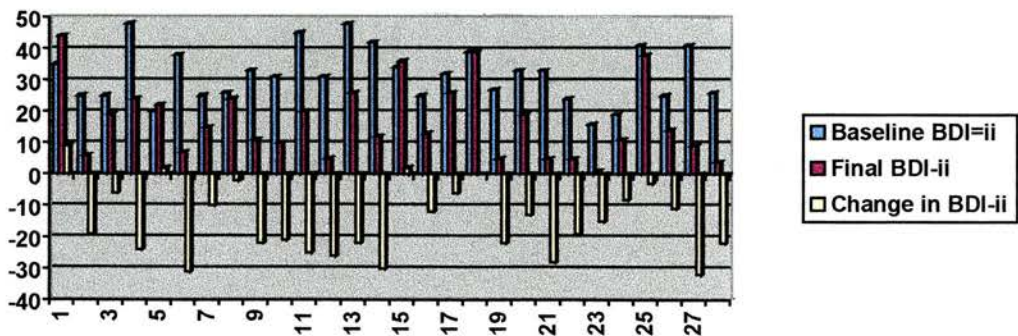
The mean BDI-II baseline, termination and change scores, for the 28 patients, are presented in Table 6. Attendance rates were high throughout the study, with patients attending a mean of 15 sessions (SD 2.3, range 6-16). Number of sessions completed was significantly correlated with end point BDI-II scores when baseline symptom severity was controlled ($r = .4$, $p = .03$), suggesting better outcome with more sessions across the patient sample. Outcome was not significantly different for first and second training cases ($t = .24$, $df = 26$, $p = .8$), nor was number of sessions attended ($t = 1.5$, $df = 10.9$, $p = .2$) or baseline symptom severity ($t = 1.2$, $df = 18.3$, $p = .24$).

Table 6: Patients' BDI-II Baseline, Termination and Change Scores

BDI-II Rating	M	SD	Range
Baseline	31.68	8.6	16 – 48
Mid Treatment	26.33	11.9	7 – 55
End of Treatment	16.79	11.8	1 – 44
Change (Baseline to Mid.)	-5.30	9.2	-24 –16
Change (Baseline to end)	-14.89	11.3	-32 – +9

At baseline 57% of the sample reported depressive symptoms within the severe range and the remaining 43.3% of subjects reported moderate severity symptoms. At end of treatment only 14% of subjects reported severe depressive symptoms, 39% reported moderate symptoms, 28% reported mild symptoms and 28% reported minimal symptoms of depression. This represents a positive change in severity classification for 75% of the sample. No patients moved from a less severe to a more severe symptom range over the course of treatment. 86% of subjects reported a reduction of symptoms from baseline to termination, with 4 subjects (20%) reporting a minimal increase in symptoms or no change, indicating that in these cases the intervention was not beneficial in reducing depressive symptoms (Figure 2). A sign test for related samples confirmed that a significantly greater proportion of subjects reduced their depressive ratings than increased them ($Z = -4.3$, $p = .001$). At termination 33% of the sample were below the BDI- II cut off for caseness (< 9).

Figure 2: Mean BDI-II Baseline, Termination and Change Scores



9.6 GROUP ADHERENCE AND COMPETENCE SCORES

The first hypothesis to be tested was that the participating therapists would be able to achieve an adequate level of adherence and competence on early casework. Previous reports employing the IPT supervision forms have employed a composite score of competence to reflect the overall quality of the therapy provided (Foley et al 1987). As previously detailed this composite score is a mean rating of all TSRF and PRF scores. To ensure that as much of the therapists work is reflected in this rating as possible it was decided to produce three competence scores to reflect the three phases of IPT. The early sessions are assessed with the initial sessions form, the middle sessions and general competence are rated on the TSRF, the PRF and the individual focus rating forms, and the final sessions are rated on the Termination form. This allows comparison of levels of competence on the different clusters of tasks and provides an indication of continuity of competence over the whole treatment. Details of the mean competence scores for the three phases are presented in Table 7.

Table 7 Mean competence scores for early, middle and termination sessions

Variable	Mean	Std. Dev.	Minimum	Maximum
INITIAL	2.66	0.62	1.60	4.25
MIDDLE/GENERAL	2.76	0.69	1.14	3.93
TERMINATION	3.24	0.82	2.14	5.33

Minimum = 7 (Very Poor); Maximum = 1 (Excellent)

This reveals that the mean competence levels for the whole group during all three phases of IPT was good to average. This was true for both initial and subsequent supervised cases. The competence ratings were not significantly different on first and subsequent cases, (Initial: $t=.6$, $df=25$, $p=.6$; Middle: $t=.6$, $df=26$, $p=.6$;

Termination: $t=1.6$, $df=23$, $p=.1$), although there was a consistent trend towards superior ratings on later casework, suggesting a non significant trend towards improved standards with additional casework (Table 8). The mean standard of practice was reasonably consistent across the three phases, although it showed a slight dip during the termination phase. This appeared to be largely the effect of first case performance. Overall the mean performance for the group was in the adequate range, supporting the competence dimension of the first hypothesis.

Table 8: Mean competence scores on first and subsequent casework

Variable	FIRST			SECOND		
	Mean	SD	Range	Mean	SD	Range
INITIAL	2.7	.65	1.60 – 4.25	2.6	0.50	1.75–3.53
MIDDLE	3.5	.94	1.14 – 3.93	2.7	1.40	1.92–3.62
FINAL	2.8	.71	1.86 – 5.33	2.6	0.62	1.97–5.00

Minimum=7 (Very Poor); Maximum=1(Excellent)

An ANCOVA, controlling for baseline symptom severity, was run to explore differences between the training groups but did not reveal significant differences on initial ($F(2,23)=.74$, ns), middle ($F(2,24)=2.56$, ns) or termination ($F(2,21)=1.1$, ns) competence scores. Baseline BDI-ii scores were not significantly correlated with any of the competence scores, suggesting that adequate competence was achieved with patients presenting across the severity spectrum, and by therapists from all of the training centres.

As incomplete data sets were available for some therapists a performance analysis was also performed using the competence rating for an early middle session for each case. The mean rating was again in the average to good range ($M=2.79$, $sd=.87$). The performance rating was not significantly correlated with the baseline severity of symptoms ($r=.13$, $p=.52$). The performance and overall competence ratings were

significantly correlated ($r = .65$, $p = .001$).

9.7 COMPETENCE OVER THE THREE PHASES OF IPT

The quality of initial sessions interventions was high (Table 9). Although the rating scale was 1 (excellent) -7 (poor), in practice the ratings were largely confined to the higher range of the scale, with, 4 = borderline, 3 = average, 2 = good and 1 = excellent. Using this system the mean rating of group performance was consistently within the average -good range, and 73.6% of individual interventions were rated as average -excellent.

Table 9: Initial Sessions Adherence and Competence Scores: Sessions 1- 4

INITIAL SESSIONS	FREQUENCY			QUALITY		
	M	SD	RANGE	M	SD	RANGE
> Symptom Review	3.1	1.0	1-4	2.6	0.9	1.3 - 5.0
> History of current episode	1.1	0.7	0-2	2.7	1.1	0.0 - 6.0
> Previous treatment	0.9	0.7	0-2	3.1	0.8	1.0 - 4.0
> Social History	0.5	0.5	0-1	3.2	0.6	2.0 - 4.0
> Explain IPT & depression	1.5	1.3	0-4	2.7	0.9	1.0 - 4.0
> Patient's expectations	0.5	0.6	0-2	2.6	0.5	2.0 - 3.0
> Symptoms in interpersonal context	2.8	1.1	0-4	2.5	0.6	1.5 - 4.0
> Positive prognosis	1.0	0.9	0-3	2.4	0.8	1.0 - 4.0
> Explain IPT techniques	0.6	0.9	0-3	2.8	0.4	2.0 - 4.0
> Negotiate contract	0.8	0.8	0-3	2.4	0.9	1.0 - 4.0
> Interpersonal inventory	2.2	1.1	0-4	2.5	0.8	1.0 - 4.0
> Formulation	1.0	0.6	0-4	2.4	0.8	1.0 - 4.0
> Treatment goals	0.2	0.4	0-1	2.8	1.0	2.0 - 4.0
> Explain respective roles	0.3	0.5	0-2	2.4	1.0	1.0 - 4.0

+ Frequency: Minimum=0, Maximum=4 -Quality: Minimum=7(Very Poor), Maximum=1(Excellent)

9.8 INITIAL SESSIONS ADHERENCE

Competence scores on the IPT forms reflect the strategies and techniques which the therapists are considered to have used. It was evident in compiling the ratings that the therapists implemented the recommended strategies with different frequencies. The pattern of strategy selection was examined to provide a better illustration of the scope and adequacy of the interventions being examined i.e. were therapists employing a sufficient range of recommended strategies to be considered adherent?

The objectives of the assessment phase of IPT are summarised in the tasks described by the initial sessions form (Table 9). This form was used to rate the interventions completed during the first four sessions. An average of 54% of initial sessions tasks were addressed in the first cases, and 67% in the second cases which were conducted. This indicates a clear increase in the percentage of tasks being addressed when subsequent casework is completed. However the range over the different cases was quite considerable, with only 21% of the specified tasks being tackled in one case, while others addressed 100%. The tasks most consistently addressed across the therapists were the review of depressive symptoms (100%), translating the depressive symptoms into the interpersonal context (96.4%) and conducting an interpersonal inventory (96.4%). These tasks are highly consistent with the principle goals of IPT, which are to address and reduce current depressive symptoms in the context of current interpersonal difficulties. The initial tasks which received least attention were the negotiation of specific treatment goals (35.7%), the negotiation of therapist and patient roles during the middle sessions (39.3%), discussion of the patient's expectations of therapy (50%) and describing the techniques employed in IPT (50%).

In order to assess which therapists met a minimal levels of adherence to the assessment tasks, the initial session strategies were reviewed and essential strategies selected in line with manual guidelines (Weissman et al, 2000). The selected strategies were symptom review, providing an explanation of IPT, translating the depressive symptoms into the interpersonal context, conducting an interpersonal inventory, providing an interpersonal formulation and establishing a contract for therapy. All essential tasks were completed in fourteen of the cases (Table 10), while between one and four of these tasks were omitted in the remaining fourteen cases. Ten of the seventeen therapists met this adherence standard on at least one case.

Two therapists met this standard on both cases submitted and two failed to meet the standard on both cases submitted. The adherent and non adherent therapists were not found to differ significantly on baseline severity of patients BDI-II score, but did differ on years of clinical experience ($t=2.5$, $df=14$, $p=.02$), with the less experienced therapists demonstrating a higher rate of adherence to the initial sessions' tasks. These adherence patterns suggest that positive adherence to the assessment tasks during the first four sessions was poor or variable for at least half of the therapists although their implementation of selected strategies was competent.

Table 10: Adherence and Competence Ratings

Therapist	Initial Sessions		Middle Sessions		Final Sessions	
	ADH	COMP	ADH	COMP	ADH	COMP
1	1,1	1,1	1,1	1,1	1,1	1,1
2	0,1	1,1	1,1	1,1	0,1	0,1
3	0,0	1,1	1,1	1,1	1,1	1,1
4	1,0,1,0	1,1,1,0	1,1,99,0	1,1,99,0	1,0,99,0	1,0,99,0
5	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	0,1,1
6	0,1	1,1	1,99	1,99	1,99	1,99
7	1	1	1	1	1	1
8	0,0	1,1	1,1	1,1	1,0	1,0
9	0	0	0	1	1	0
10	0	1	1	1	1	1
11	0	1	1	1	0	0
12	1	1	1	1	1	1
13	1	1	1	0	0	0
14	0,1	1,1	1,1	1,1	1,1	0,1
15	0	1	0	0	0	0
16	0	0	1	0	0	0
17	1	1	0	0	0	0

Each figure represents a supervised case. 1 = meets criteria; 0=does not meet criteria; 99=missing data

9.9 MIDDLE SESSIONS ADHERENCE

During the middle sessions specific strategies are outlined for each of the focus areas. Table 11 outlines the focus areas negotiated in the individual casework. Identifying a focus area is a central strategy in IPT and was completed in all cases, with explicit negotiation during the first four sessions in twenty five cases, and

performance ranging from average to excellent. Interpersonal role transitions was the most commonly selected focus, accounting for eighteen cases (64.3%), followed by interpersonal disputes which was the choice in eight (28.6%). Only two cases (7.1%) were specifically formulated in terms unresolved grief, and consistent with the recommendations of the IPT authors that interpersonal deficits be used infrequently, this was not selected. The selection of foci is consistent with Wolfson et al's (1997) study of IPT foci in depressed elders in which the rank order was the same with 41 %, 34.5%, 23% and 1.5% respectively for the four areas. The focus groups did not differ significantly on baseline symptom severity ($F(2,25)=.48, p=.6$).

Table 11: IPT Focus Selection

Therapist	Role Disputes	Role Transition	Grief	Interpersonal Deficits
1	0	2	0	0
2	0	2	0	0
3	1	0	1	0
4	1	3	0	0
5	1	2	0	0
6	1	1	0	0
7	1	0	0	0
8	1	1	0	0
9	0	1	0	0
10	1	0	0	0
11	0	1	0	0
12	0	1	0	0
13	0	1	0	0
14	0	2	0	0
15	1	0	0	0
16	0	0	1	0
17	0	1	0	0
TOTAL	8(28.6%)	18 (64.3%)	2 (7.1%)	0 (0%)

9.10 TSRF

During the middle phase, ratings were made on the TSRF, PRF and individual focus forms. Although each of the TSRF strategies may be appropriately employed within any of the individual cases, certain strategies are more characteristic of each of the foci, and may be expected to predominate according to choice. A series of ANOVAs were used to compare the mean frequency and quality of the individual strategies

across the focus groups. These revealed a number of significant differences between the focus groups on the frequency with which the individual strategies were employed but no significant differences in the competence in applying them. This suggests that therapists were able to appropriately select the strategies consistent with their negotiated focus and to implement the range competently. Details are presented in Table 12.

Table 12: ANOVAs: TSRF ADHERENCE AND COMPETENCE RATINGS BY FOCUS

		FREQUENCY			QUALITY		
		F	df	Sig.	F	df	Sig.
Explore recent and remote losses	Between Groups	12.73	2	.01*	0.22	2	.81
	Within Groups		23			12	
	TOTAL		25			14	
Facilitate Mourning	Between Groups	146.33	2	.01*	10.76	1	.08
	Within Groups		23			2	
	TOTAL		25			3	
Resume/Develop Relationships	Between Groups	0.84	2	.44	0.40	2	.67
	Within Groups		23			21	
	TOTAL		25			23	
Nature of Dispute/Transition	Between Groups	8.20	2	.01*	0.43	2	.66
	Within Groups		23			21	
	TOTAL		25			23	
Patient's position Dispute/Transition	Between Groups	5.73	2	.01*	0.65	2	.53
	Within Groups		23			22	
	TOTAL		25			24	
Possible Changes	Between Groups	3.27	2	.06	0.07	2	.94
	Within Groups		23			23	
	TOTAL		25			25	
Past/Current Relationships	Between Groups	0.04	2	.96	0.10	2	.90
	Within Groups		23			10	
	TOTAL		25			12	
Review of Self Concept	Between Groups	2.15	2	.14	1.77	2	.28
	Within Groups		23			4	
	TOTAL		25			6	
Patient/Therapist Relationship	Between Groups	0.63	2	.54	-	-	-
	Within Groups		23				
	TOTAL		25				

In reviewing the mean frequencies of the strategies for each focus group (Table 13) it was evident that the grief cases used the loss and mourning strategies more regularly than the other two focus groups, which used them very infrequently. Resuming and developing relationships continued to be used almost twice as frequently in the grief cases, but received greater attention in the other cases and therefore was only significant at a trend level. The two central transition and disputes strategies were both used frequently in the corresponding cases but very infrequently in the grief cases. It was of note that the disputes cases consistently demonstrated higher mean frequencies than transitions, suggesting more strategy based work in these cases. The exploration of possible change approached a significant difference across the groups, reflecting the fact that it was used three to four times as often in the transition and disputes as in the grief cases. The self concept and the therapeutic relationship were infrequently the focus on any of the cases. These would be most consistent with a deficits focus or in addressing ruptures in the therapeutic alliance, and it was therefore consistent that these were selected infrequently. This demonstrates that each focus group adhered to the strategies most consistent with the contracted area of interpersonal difficulty, and made minimal to no use of other strategies.

Table 13: Focus Groups: TSRF Strategy Application

TSRF STRATEGIES	Focus	N	Mean	SD	Min.	Max.
Explore recent and remote losses	Grief	2	11.0	2.8	0.0	13.0
	Transition	15	1.7	2.6	0.0	8.0
	Disputes	8	1.6	2.4	0.0	7.0
	TOTAL	25	2.4	3.6	0.0	13.0
Facilitate Mourning	Grief	2	6.5	2.1	5.0	8.0
	Transition	15	0.0	0.0	0.0	0.0
	Disputes	8	0.3	0.5	0.0	1.0
	TOTAL	25	0.6	1.8	0.0	8.0
Resume/Develop Relationships	Grief	2	7.5	7.8	2.0	13.0
	Transition	15	4.0	3.6	0.0	12.0
	Disputes	8	4.5	3.8	0.0	9.0
	TOTAL	25	4.4	3.7	0.0	13.0
Nature of Dispute/Transition	Grief	2	0.5	0.7	0.0	1.0
	Transition	15	5.0	2.0	4.0	9.0
	Disputes	8	8.4	3.8	0.0	15.0
	TOTAL	25	5.7	3.4	0.0	15.0
Patient's position in Dispute/Transition	Grief	2	1.5	2.1	2.0	3.0
	Transition	15	9.1	3.0	7.0	13.0
	Disputes	8	10.4	3.5	0.0	16.0
	TOTAL	25	8.9	3.8	1.0	16.0
Possible Changes	Grief	2	2.0	1.4	1.0	3.0
	Transition	15	8.2	3.3	6.0	12.0
	Disputes	8	8.6	3.1	1.0	15.0
	TOTAL	25	7.8	3.5	0.0	15.0
Past/Current Relationships	Grief	2	2.0	1.4	0.0	3.0
	Transition	15	1.4	2.6	0.0	7.0
	Disputes	8	1.5	1.8	0.0	5.0
	TOTAL	25	1.5	2.2	0.0	7.0
Review of Self Concept	Grief	2	2.0	2.2	0.0	4.0
	Transition	15	0.3	0.6	0.0	2.0
	Disputes	8	0.8	1.5	0.0	4.0
	TOTAL	25	0.6	1.2	0.0	4.0
Patient/Therapist Relationship	Grief	2	0.0	0.0	0.0	0.0
	Transition	15	0.1	0.6	0.0	1.0
	Disputes	8	0.0	0.0	0.0	0.0
	TOTAL	25	0.0	0.3	0.0	1.0

In all of the transition and disputes cases, except two, there was evidence of all of the appropriate strategies being employed. In both of the grief cases each of the central strategies were employed. It was of note that the two cases in which a central strategy was omitted were also the two cases in which the highest number of initial sessions tasks were omitted, suggesting a generally low pattern of adherence for these cases.

The frequency with which the strategies were employed was quite variable across therapists. As has been noted there are no specific guidelines on the frequency with which individual strategies may be optimally employed, and a fifty percent cut off

point was employed to identify those examples in which the appropriate strategies were implemented more often than not. Eleven of the twenty six (42%) transition and disputes cases were rated as employing at least three of the appropriate strategies for the majority of sessions. Both grief cases employed the appropriate strategies during the majority of sessions. This reflected the casework of ten of the therapists, with four meeting criteria on two consecutive cases. Six of these therapists had also met adherence criteria on initial sessions tasks. It should be noted that while one therapist consistently employed the appropriate strategies the mean competence rating on each was below an acceptable level.

9.11 PRF

The items on the PRF are equally applicable in each phase and each focus area of IPT. The character of these interventions was demonstrated to be largely non-directive, with repeated attention given to the affective experience and developing and clarifying interpersonal awareness. The use of techniques inconsistent with the IPT manual guidelines, as described by Klerman et al (1984) e.g. psychoanalytic interpretations, specific cognitive interventions, was minimal, although not absent. The most commonly used strategies were exploratory techniques, exploring affect and clarification, which were all evident in the majority of sessions for twenty three of the twenty five cases with seven or more PRF ratings. The two exceptions used exploration of affect and clarification in only a minority of sessions. One of these therapists had failed to meet previously reported adherence standards. Communication analysis and directive techniques were used more intermittently and reference to the therapeutic relationship, use of a significant other and decision analysis were used infrequently. Once again the mean quality of the interventions were average to good for all of these techniques. ANOVAs revealed no significant

differences between the focus groups in number of process strategies used ($F(2,21)=.3$, $p=.74$) or in overall quality of interventions ($F(2,21)=.77$, $p=.47$). Details of frequency and quality ratings are presented in Table 14.

Table 14: Group Mean PRF Adherence and Competence Scores

PROCESS STRATEGIES		Frequency			Quality		
		M	SD	Range	M	SD	Range
➤	Exploratory Techniques	12.5	2.7	7 -16	2.4	0.8	1.3 – 3.8
➤	Admin. Details	2.5	2.1	0 - 6	2.5	2.1	0.0 – 6.0
➤	Explore Affect	10.9	3.6	3 -16	2.7	0.6	1.3 – 3.8
➤	Clarification	11.2	3.8	0 -16	2.4	0.7	1.4 – 3.8
➤	Communication Analysis	1.8	2.2	0 - 9	2.7	0.7	1.3 – 4.0
➤	Therapeutic Relationship	0.2	0.4	0 -1	3.0	1.0	2.0 – 4.0
➤	Directive Techniques	6.6	3.7	1 -14	2.2	0.4	1.7 – 3.0
➤	Decision Analysis	2.0	0.6	0 -2	3.0	0.0	3.0 – 3.0
➤	Significant Other	0.1	0.3	0 -1	3.0	0.0	3.0 – 3.0
➤	Non IPT	1.3	1.1	0 - 3	-	-	-

Frequency: Minimum=0, Maximum = 16, Quality: Minimum=7, Maximum=1

9.12 FOCUS SPECIFIC STRATEGIES

The two interventions common to each focus area are symptom review and linking symptom onset to the focus area. These interventions are the suggested starting point for each IPT session. The symptom review was evident in the majority of sessions for twenty one of the twenty five cases reviewed (84%), while the link to focus was evident in the majority of sessions for eighteen of these cases (72%). Cases were omitted if there was insufficient data to reflect the majority of the middle sessions work. ANOVAs did not reveal a significant difference in the frequency or quality of use of these two strategies between the focus groups. The difference in frequency of use of the two strategies may reflect a general tendency for some therapists to address the symptoms of depression and the areas of interpersonal difficulty in sequence and to fail to make connections between the two. Details for the three focus areas employed are presented in Tables 15-17.

Table 15: Role Transition Frequency and Competence Scores (n=18)

STRATEGY	Frequency		Min	Max	Quality		Min	Max
	M	SD			M	SD		
➤ Review symptoms	8.7	2.9	1	12	2.8	0.7	1.7	3.8
➤ Relate symptoms to focus	6.9	3.1	2	13	3.0	0.9	1.8	4.1
➤ +/- of old and new roles	4.0	2.2	0	8	2.8	0.7	2.0	4.0
➤ Feelings re what is lost	2.4	2.2	0	9	3.0	0.6	2.0	4.3
➤ Feelings re the change	1.5	1.5	0	4	2.9	0.7	2.0	4.0
➤ Opportunities in new role	5.4	2.8	1	10	2.8	0.7	1.8	3.8
➤ Evaluate loss	1.2	1.9	0	7	3.0	0.5	2.3	4.0
➤ Release of affect	6.9	4.0	0	12	2.8	0.6	1.9	3.6
➤ Develop social support	7.3	3.3	1	12	2.7	0.8	1.5	4.0

+ Frequency: Minimum=0, Maximum=12 -Quality: Minimum=7(Very Poor), Maximum=1(Excellent)

Two therapists demonstrated inadequate evidence of either of the symptom focused tasks, and five additional therapists failed to perform one of the tasks during the majority of sessions. Both therapists who failed to consistently focus on depressive symptoms and their link to the focus had failed to meet most of the previous adherence criteria (Table 10).

Table 16: Disputes Frequency and Competence Scores (n=8)

STRATEGY	Frequency		Min.	Max.	Quality		Min.	Max.
	M	SD			M	SD		
Review symptoms	7.5	2.6	4	12	3.0	0.8	2.1	4.6
Relate symptoms to focus	7.3	2.5	4	12	2.8	0.9	1.6	4.8
Stage Dispute	3.0	1.9	1	7	2.8	1.1	1.4	4.5
Dispute Issues	5.4	2.4	2	10	2.6	0.9	1.5	4
Non Reciprocal Role Expectations	2.3	3.1	0	8	2.1	0.3	1.8	2.5
Differences in values	4.4	2.2	2	9	2.6	0.6	1.8	3.7
Relationship Parallels	3.1	2.4	0	7	2.9	1.0	1.6	4.0
Options for change	7.4	1.6	5	10	2.7	1.0	1.5	4.6
Communication Patterns	5.1	1.7	3	8	2.6	1.0	1.7	4.8
How is dispute perpetuated	2.7	3.1	0	7	2.7	1.0	1.5	4

+ Frequency: Minimum=0, Maximum=12 -Quality: Minimum=7(Very Poor), Maximum=1 (Excellent)

Table 17: Grief Frequency and Competence Scores (n=2)

Strategies	Frequency				Quality			
	M	SD	Min	Max	M	SD	Min	Max
Review Symptoms	9.0	1.4	8	10	3.3	0.7	2.8	3.8
Relate symptoms to focus	8.0	2.8	6	10	3.2	0.4	2.9	3.5
Reconstruct Relationship	4.0	1.4	3	5	2.7	1.9	1.3	4.0
Events around death	2.5	2.1	1	4	2.6	2.3	1.0	4.3
Social support around mourning	7.0	1.4	6	8	2.9	1.3	2.0	3.8
Explore associated feelings	5.5	2.1	4	7	2.6	1.6	1.5	3.7
Involved with others	6.5	4.9	3	10	2.5	1.1	1.7	3.3

+ Frequency: Minimum=0, Maximum=12 -Quality: Minimum=7(very Poor), Maximum=1 (Excellent)

9.13 TERMINATION SESSION ADHERENCE

Mean competence scores on application of the Termination sessions strategies were found to be slightly lower than those of the earlier phases of therapy. This may at least in part be attributable to the minimal attention this area was given by many of the therapists. Two patients ended their contact abruptly and without prior warning and these cases are not evaluated on preparation for the end of therapy, as it did not happen. Five further cases addressed the end of therapy only in the final session. In four of these cases this was also connected with the early conclusion of therapy, typically around session twelve, but in these cases this was discussed with the therapist in advance. In two of these cases the early conclusion of therapy was the therapist's responsibility. Once because the session numbers had not been monitored and the contracted time ran out and once was because the therapist assessed the patient as having improved sufficiently and suggested that the sessions conclude before the contracted time. Neither strategy is consistent with the adherent or competent application of the IPT manual guidelines. In the final case in which minimal attention was given to the termination of therapy all sixteen session were conducted

but the ending received little to no attention. Frequency and quality scores are presented in Table 18.

Table 18: Termination Sessions Frequency and Competence Scores

TERMINATION STRATEGIES	FREQUENCY			QUALITY		
	M	SD	RANGE	M	SD	RANGE
Discuss ending	2.4	1.3	0 – 4	2.2	1.0	1.0 – 4.0
Reaction to ending	1.9	1.1	0 – 4	2.1	0.7	1.0 – 3.3
Grief at ending	1.0	1.0	0 – 3	2.0	0.8	1.0 – 3.0
Independent Competence	2.7	1.5	0 – 4	2.4	0.8	1.0 – 4.0
Review Treatment	1.3	1.0	0 – 3	2.3	1.0	1.0 – 4.0
Evaluate Treatment	1.4	1.4	0 – 4	2.5	0.8	1.0 – 4.0
Early Warnings	0.9	1.3	0 – 4	2.6	0.7	2.0 – 4.0

Frequency: Minimum=0, Maximum=4 Quality: Minimum=7 (Very Poor), Maximum=1 (Excellent)

The IPT manual guides the therapist to an explicit and reflective end to therapy. One means of assessing adherence to this would therefore be to look at how many therapists directly spoke to their patients about the end of contact. The end of therapy was mentioned in twenty three (88%) of the cases in which patients did not default from treatment. In twenty two (84%) of these cases the patients reaction to the end of therapy was also discussed. It should be noted however that in eight of the cases in which the end of therapy was discussed it was rated as having been done at a below average level of competence. In five of these cases the facilitation offered for the patients in exploring their reaction to the end of treatment was also rated as below average. It was also of note that two of the therapists who had failed to meet previous adherence and competence criteria were again represented in this sample. It would appear then that approximately one third of therapists failed to meet the basic adherence criteria of explicit and adequate discussion of the end of therapy.

In summary, nine (32%) of the cases reviewed were judged to have met all adherence and competence criteria. This reflected the work of seven of the therapists (41%). Performance was variable across cases, and only three therapists consistently met all of the adherence and competence standards. The early and late sessions were most vulnerable to omission of central strategies or incompetent performance. When the general adherence and competence after formulation was reviewed eighteen cases (64%) met the reported standards, reflecting the work of eleven of the therapists (65%).

9.14 INTER RATER RELIABILITY

In order to assess the reliability of the ratings made during supervision, two experienced IPT supervisors were asked to make second ratings of a selection of tapes. Individual tapes were selected from early in the middle phase, mainly sessions five or six, and fourteen of the seventeen therapists were represented in the second rating sample. Given the timing of the rating the PRF, TSRF and appropriate Focus forms were second rated. The second raters were blind to the therapists' identities and to treatment outcome. They were informed of the session they were rating, to provide the contextual information to make their assessment, but were not informed of the selected focus until after the tape had been reviewed. This information was provided prior to making the ratings to ensure the focus rating forms corresponded.

Ratings on each of the forms were reviewed for correspondence in adherence scores i.e. presence and absence of each intervention, and competence scores. Correspondence on competence scores was noted in four ways – exact correspondence, scores within one point of each other, correspondence on pass/fail,

and different scores i.e. two or more points apart on the seven point scale. Details of the correspondence between adherence ratings are presented in Table 19. This demonstrates a highly consistent and good level of correspondence across the different forms, with the clear majority of items being consistently rated as absent or present by different raters.

Table 19: Percentage agreement on PRF, TSRF and Focus Forms Adherence Scores

Form	% of Exact Agreement	KAPPA
PRF	84%	.66
TSRF	78%	.55
Transitions	79%	.55
Disputes	80%	.58
Grief	79%	.43

A further and more statistically robust check on the reliability of these dichotomous variables was provided by calculating Kappa scores (Table 19), to control for chance correspondence between scores. This generated Kappa ratings which reflected substantial inter rater agreement on the PRF and moderate agreement between raters on the TSRF and on each of the focus forms employed.

When the competence scores were compared an initial high standard of exact agreement was used. It became apparent however that many of the ratings were similar to each other but did not match exactly and so a second calculation was made to represent the extent to which the different raters produced similar but not matching scores i.e. one point apart on the scale. Given that summary competence scores have also been employed in many of the analyses it also seemed appropriate to

examine the correspondence of these scores in rating the therapists on overall categorisation of competence or incompetent practice (Table 20).

It was evident that the majority of ratings were clustered around one end of the rating scale. This did not produce a normal distribution of scores, and so Kendall's Tau, a nonparametric measure of correlation, was calculated. Kappa could not be used as this is calculated for nominal data, while the competence ratings were measured on an interval scale. Overall correspondence on focus ratings was calculated rather than individual foci scores in order to maximise the sample size. Only two of the ratings, TSRF and the composite Middle sessions rating, demonstrated a significant correlation.

Table 20: Percentage agreement, near agreement and discrepancy on PRF, TSRF and Focus Form Competence Scores

Form	% Exact Agreement	% Agreement plus +/- 1	% Adequate/Inadequate	% Disagreement	KENDALL'S TAU (p)
PRF	55%	74%	79%	26%	-.21 (ns)
TSRF	59%	72%	86%	28%	.39 (p=.03)
Transitions	38%	70%	86%	30%	-
Disputes	54%	68%	100%	32%	-
Grief	50%	79%	50%	21%	-
Focus	-	-	-	-	-.12 (ns)
Middle	-	86%	86%	14%	.73 (p> .001)

It is evident that correspondence between the competence ratings on individual items was not as high as that demonstrated on the adherence scores. When exact agreement is the standard, a lower but adequate level of agreement was achieved on all forms except for focus on transition cases. This level was substantially improved when the less rigorous standard of agreement to within one point of the first rating

was employed, suggesting a good level of general agreement among raters on the quality of the work being reviewed. The overall categorisation of practice as competent or not was highly consistent between raters on most scales. This general correspondence is important when considering the less consistent correlation ratings. Within the PRF and focus scales the raters, while largely clustered within one point of each other, were as likely to score one point higher or one point lower than each other, resulting in a poor pattern of correlation. Across subjects however the pattern was highly consistent, with the pattern of higher or lower rating between raters being reversed on only one of the three scales for two of the tapes reviewed.

9.15 COMPETENCE AND OUTCOME

The third hypothesis was that higher levels of competence in performing IPT would be related to greater reductions in BDI-ii scores by the end of treatment. Pearson's r was calculated between the competence scores for the middle focus sessions and both the end point BDI-ii scores and the change in BDI-ii scores, for those cases in which more than half of the sessions were rated. Neither the end point ($r=.03$, $p=.87$) nor the change score ($r=.01$, $p=.94$) were significantly correlated with the competence rating for therapy. While this may suggest that level of competence and reduction in symptomatic distress are unrelated in this sample it should also be noted that there was a generally high level of competence across therapists, and therefore there may have been insufficient variance to reveal an association.

A second set of correlations were run between the performance rating for the early middle session and end of treatment and change scores on the BDI-ii. This revealed a significant relationship between the performance rating and the change in symptom rating ($r=-.38$, $p=.05$), indicating greater change with more competent practice. This association was no longer significant when baseline symptoms severity and number

of sessions attended were controlled for ($r = -.24$, $p = .24$). End point BDI-II and performance rating were not significantly correlated.

9.16 PERFORMANCE, OUTCOME AND THERAPEUTIC ALLEGIANCE

As therapist allegiance to a treatment model has been shown to account for a greater proportion of outcome variance than differences in treatment approach (Robinson et al, 1990), therapists were asked to indicate their preferred treatment model for major depression. Thirteen therapists reported more than one preferred approach, suggesting a continuum rather than a dichotomous perspective. Twelve therapists noted IPT among their preferred models of treatment, six placing IPT ahead of CBT, one placing it ahead of a psychodynamic approach, four placing it behind CBT, and one placing it behind emotion focused cognitive therapy. Two therapists described a preference for an eclectic approach, indicating CBT and psychodynamic influences specifically, one reported CBT to be their preferred model and one reported goal setting/problem solving. These results suggest that the therapists were IPT friendly, although not exclusive advocates of this approach.

Although IPT cannot be compared to another treatment model, reports of preference were used to explore the impact of therapeutic preference on treatment outcome. Given the small numbers reporting preference for individual models, the therapists were split according to whether they reported their preferred model to be IPT or not. Roughly half the therapists were in each group, and therapeutic preference did not reveal a significant difference in treatment outcome defined as end point score ($t =$

1.6, $df=26$, $p=.1$) and BDI-II change score ($t=1.4$, $df=26$, $p=.2$). The mean end of treatment score for patients of IPT advocates was 20.8 ($SD=13.4$), while it was 13.75 ($SD=9.8$) for therapists who preferred an alternative therapy model. It was of interest to note that despite almost identical baseline BDI-II scores the group who did not report that IPT was their preferred model produced greater improvement on the BDI-II over the course of treatment.

The therapists who had identified IPT as their first choice and those who had not were also compared on their competence ratings. This revealed no difference on competence ratings for initial sessions ($t=.3$, $df=13.8$, $p=.78$), focus sessions ($t=.4$, $df=26$, $p=.69$) or termination sessions ($t=.2$, $df=12.1$, $p=.88$). This confirms that ratings of therapists' performance and patient outcome did not reflect the therapists' reported therapeutic allegiance.

9.17 EXPERIENCE AND OUTCOME

When outcome was analysed with reference to therapist experience, mixed evidence emerged on the influence of clinical experience on treatment outcome. Experience was defined in a number of ways: years of clinical practice, number of depression cases treated, and type of clinical qualification (Table 21). Outcome was defined both as termination BDI-II score and as change in BDI-II scores.

Correlations between years of clinical practice and BDI-II change ($r=.21$, $p=.29$) and BDI-II termination scores ($r=.21$, $p=.29$), controlled for initial symptom severity, were

not significant. The remaining data on experience was nominal, and in order to prevent the numbers in each group becoming virtually equivalent to comparing individuals, data for the two remaining definitions of experience were re-classified. Clinical qualification was used to reflect a psychological or psychiatric training background or the absence of a professional qualification. This split was employed because Blatt et al (1996) reported a significantly higher percentage of clinical psychologists than psychiatrists in their more effective therapist group. The number of

Table 21:Therapists' Experience

Therapist	Years of Clinical Practice	No. of Depression Cases Treated	Type of Clinical Qualification
1	3	11-20	Psychology
2	2	21+	Psychology
3	19	21+	Psychology
4	6	21+	Psychology
5	4	0-10	None
6	4	21+	None
7	1	0-10	Psychiatry
8	20	21+	Psychology
9	15	21+	Psychiatry
10	2	11-20	Psychology
11	10	21+	Psychiatry
12	1	0-11	None
13	15	21+	Psychiatry
14	14	21+	Psychiatry
15	8	21+	Psychiatry
16	10	21+	Psychiatry
17	10	21+	Psychiatry

patients with depression previously treated was dichotomised to above and below 20, as all therapists above the threshold were considerably so. Baseline symptom severity was not significantly different between groups when experience was defined in terms of either, number of patients with depression treated ($t=.63$, $df=26$, $p=.54$) or type of clinical training ($F(2,25)=2.6$, $p=.1$). Independent t-tests and ANOVAs were run for each comparison between level of experience and outcome. No significant differences emerged on either end of treatment scores for groups defined by number

of patients with depression previously treated ($t=.18$, $df=26$, $p=.85$) or type of clinical qualification ($F(2,25)=.49$, $p=.62$) or on change in symptoms severity for the respective groups ($t=.67$, $df=26$, $p=.5$; $F(2,25)=3.06$, $p=.065$). Change in BDI-ii scores had approached significance and post hoc comparisons (LSD test) revealed significantly ($p=.02$) better outcome for those therapists without a clinical qualification compared to those with a psychiatry qualification, but not compared to those with psychology training. More conservative post hoc comparisons (Scheffe test) did not replicate this significant finding although there was a trend towards significance between the unqualified and the psychiatric group ($p=.07$). These findings largely support the proposal that treatment outcome would not be significantly influenced by the therapists' clinical experience, although there was some modest evidence for the superiority of unqualified therapists over psychiatry trained therapists.

9.18 EXPERIENCE AND PERFORMANCE

Therapist performance was also reviewed with reference to clinical experience. In the first instance years of clinical experience was correlated with competence scores. As length of experience has been shown to correlate with baseline BDI-ii scores, all competence and experience correlations were controlled for baseline symptom severity. Initial ($r=.2$, $p=.4$), middle ($r=.1$, $p=.5$) and termination ($r=.1$, $p=.6$) competence scores were not significantly correlated with years of clinical experience.

Therapists' experience with depressed patients was used in the next analysis. Again none of the comparisons were significant. This may reflect the cut off which was required in order to divide the sample i.e. having seen more or less than twenty depressed patients. The sample had generally seen a high number of patients with

depression and only three of the seventeen therapists were in the less experienced group, which is an insufficient number to explore the relationship adequately. Information on experience with depressed patients was collected in a categorical form, as it seemed unlikely that therapists would be able to provide an exact figure and therefore correlations could not be run to explore the general pattern of association.

When type of clinical experience and training was used to look at competence ratings a series of significant differences emerged. Significant differences were revealed on all the competence scores, with the exception of competence on the initial sessions tasks (Table 22). Post hoc Scheffe tests revealed that psychiatry trained therapists were significantly less competent than psychology trained therapists during the middle ($p=.03$) and termination ($p>.01$) phases of treatment, and unqualified therapists were not significantly different from either. This pattern was evident on each of the middle phase assessments – PRF ($p>.01$), TSRF ($p=.05$) and Focus ($p>.01$).

Table 22: ANOVA: Competence ratings and training background

Competence Ratings	F	Df	p
Initial Sessions	1.1	2	.35
TSRF	3.4	2	.05
PRF	6.6	2	.01
Focus	9.9	2	.01
Middle	4.4	2	.03
Termination	6.2	2	.01

Review of mean scores demonstrated that the therapists with a psychology background were consistently rated highest, those with a psychiatry background consistently received the poorest ratings and those therapists without a clinical qualification were rated between the other two groups (Table 22). Given that three

therapists without a clinical qualification all held psychology degrees and were working as psychology/research assistants to qualified clinical psychologists their scores were collapsed with the other psychology scores and the two groups were compared again. This did not alter the pattern of results.

Table 23: Mean (SD) Competence ratings by training background

Competence Ratings	PSYCHOLOGY		PSYCHIATRY		UNQUALIFIED	
	M	SD	M	SD	M	SD
Initial Sessions	2.5	.6	2.9	.8	2.7	.2
TSRF	2.4	.6	3.1	.6	2.7	.4
PRF	2.3	.5	3.1	.6	2.7	.3
Focus	2.4	.5	3.5	.6	2.9	.4
Middle	2.4	.5	3.2	.6	2.6	.7
Termination	2.6	1.2	3.9	.9	3.1	.5

Thus the hypothesis that competence and experience are independent is partially rejected. Definition of experience strongly influences the results such that years of clinical experience and number of patients with depression treated does not relate to competence on training cases, while psychology training is related to significantly higher competence than psychiatry training.

9.19 INITIAL SYMPTOM SEVERITY AND OUTCOME

Previous reports on IPT (Elkin et al,1985) found that IPT worked well as a treatment for both moderately and severely depressed patients. This would predict a consistent level of clinical change for patients across the severity spectrum. The sample was divided according to baseline severity, using thirty on the BDI-ii as the cutoff point for

the severe group. In order to gauge the clinical response to treatment the BDI-ii change score, reduction in severity rating and meeting recovery criterion ($\text{BDI-ii} < 9$) were used as the outcome measures. End of treatment score would not be suitable, as even with equivalent symptom change the two groups would be different at the end of treatment given the baseline disparity. The equivalence of response in the two groups was supported by the non significant difference revealed for the BDI-ii change score ($t=1.2$, $df=26$, $p=.24$), indicating that severely and moderately depressed patients reported the same amount of gain on the BDI-ii across treatment.

When treatment response was defined as a drop in severity range over the course of treatment twelve (80%) of the severely depressed patients met this criteria. Chi square analysis did not reveal a significant difference between the severe and moderates groups in number of patients reducing symptom severity level over treatment.

A Mann-Whitney test was run to compare number of severely and moderately depressed patients meeting recovery criteria and this revealed the two groups did not differ in the number who met this goal ($z=.65$, $p=.5$), supporting the hypothesis that severely and moderately depressed patients demonstrate an equivalent response to IPT.

9.20 EARLY SYMPTOM CHANGE AND FINAL OUTCOME

BDI-ii ratings were available for the majority of patients at an early middle session. This revealed the early symptom response experienced. The change in BDI-ii rating early in treatment was correlated with the end of treatment rating, with baseline severity controlled, and this revealed a significant relationship ($r=.7$, $p=.001$). When the sample was split according to recovered and not recovered at the end of

treatment, the two groups were significantly different on early treatment symptom change ($t= 2.17$, $df=25$, $p=.04$), and the same was shown for patients reducing severity rating ($t=3.4$, $df=19$, $p=.003$), indicating that greater early change was related to better final outcome status.

9.21 PROBLEMS IN CONDUCTING THERAPY

During the course of therapy, ratings were made to reflect problems in conducting therapy. The options for recording problems spanned a number of factors including missed appointments, lateness, uncooperativeness in the session, suicide threats or attempts, impersonal presentation and early termination of therapy. A global rating was calculated to represent the total number of problems noted within individual cases across the whole contact. An initial correlation was run with baseline symptom severity to establish whether the problems noted in therapy were independent of the severity of the depression being treated. This was confirmed by the non significant relationship which was revealed ($r=.006$, $p=.98$). The global rating was then correlated with the end of treatment BDI-ii ($r=.49$, $p=.009$), change in BDI-ii rating achieved from baseline to end of treatment ($r=.51$, $p=.006$), and mid treatment BDI-ii ($r=.62$, $p=.001$). All revealed highly significant correlations, demonstrating that the greater the number of problems noted during therapy the worse the clinical outcome both during and after completion of treatment. This finding was further supported by the significant difference identified between those subjects who achieved the pre determined recovery standard on the BDI-ii and those who did not in terms of problems in therapy ($t=2.14$, $df=26$, $p=.04$), with those who recovered rated with less problems in therapy.

The global rating was then correlated with the competence ratings. This did not reveal a significant relationship between the overall competence rating and the global

problem rating, but did reveal a relationship between early competence ratings and overall problems in conducting therapy. Both the initial sessions competence rating ($r=.65$, $p=.001$) and the performance rating early in the middle session ($r=.399$, $p=.04$) were significantly related to the overall rating of problems. This suggests that the noted problems were associated with poorer performance in the earlier phases of therapy, but were not related to competence levels in the middle and final phases of treatment.

9.22 MOST AND LEAST EFFECTIVE THERAPISTS

Therapist efficacy was defined in terms of the mean end of treatment scores and change in symptom severity reported at the end of therapy. A mean score was calculated for each therapist's casework and the four therapists with the lowest end point scores and greatest change in symptom severity were compared with the four therapists with the opposite pattern to explore the factors which may have contributed to this treatment effect. An initial comparison confirmed that the difference in final outcome was not simply a reflection of initial symptom severity ($t=.0$, $df=8$, $p=1.0$), which was almost identical for the two groups.

Practical factors reflecting timing and duration were considered next, but no significant differences were found in relation to delay to conducting supervised casework ($t=.69$, $df=8$, $p=.5$), number of supervision sessions attended ($t=.14$, $df=8$, $p=.89$) or number of therapy sessions conducted ($t=1.08$, $df=6.47$, $p=.32$). The trend in the data indicated that the therapists with worst outcome waited longer for casework to begin ($M=14.3$ mo. ($SD=11.4$) v 9.7 mo. ($SD=7.8$), attended fewer supervision sessions ($M=14.6$, $SD=4.5$ v $M=15$, $SD=2.1$) and conducted fewer therapy sessions ($M=15.5$, $SD=.5$ v $M=14.6$, $SD=1.7$). None of the therapist defined variables i.e. duration of clinical experience ($t=2$, $df=7.8$, $p=.08$), preferred treatment

model for depression ($Z=-1.75$, $p=.17$), training background ($Z=-1.4$, $p=.25$), experience with depressed patients ($Z=-.27$, $p=.9$), were significantly different for the most and least effective therapists. The pattern in the data suggested that the therapists with worse outcome had more clinical experience, were more likely to express a therapeutic preference in favour of IPT and to have had a psychiatric training.

When in therapy factors were considered only one significant difference emerged. The global rating of problems conducting therapy ($t=1.9$, $df=8$, $p=.08$) approached significance, with greater problems evident in the casework of the least effective therapists, while change in BDI-ii rating early in therapy ($t= 4.6$, $df=7$, $p=.002$) was significant. Greater early reduction rather than increase in symptom severity ($M=-14.5$, $SD=5.5$ v $M=4.6$, $SD=6.6$) characterised the work of the most effective therapists. None of the competence scores, for initial ($t=1.8$, $df=8$, $p=.85$) or termination phases ($t=7.2$, $df=7$, $p=.5$), individual sessions ($t=1.5$, $df=8$, $p=1.6$) or for general competence ($t=1.8$, $df=8$, $p=.85$), were significantly different for therapists with the best and worst outcome, and all mean scores fell within the acceptable range.

10. DISCUSSION

10.1 ADHERENCE AND COMPETENCE

Having undertaken a program of training and supervision in Interpersonal Psychotherapy a group of therapists, with a wide range of clinical experience, and from a variety of theoretical backgrounds, performed a high proportion of the prescribed tasks to an average to good standard. IPT progresses through three stages, initial, middle and late, with a distinct pattern of strategies characterising the interventions at each stage. Therapists' mean competence ratings remained in the average to good range across the three stages of the intervention, and from first training case to second. This demonstrated that therapists in general were able to conduct an assessment, implement an intervention with reference to an area of interpersonal difficulty, and conclude the therapeutic contact with comparable and acceptable levels of competence across cases.

Detailed review of the strategies selected by the therapists however suggested that many of the therapists failed to meet at least some of the adherence criteria proposed for each phase of IPT i.e. completion of central tasks within the recommended stage of therapy. Only three therapists were rated as meeting all the criteria on all submitted casework, while four additional therapists met criteria on at least one supervised case. The assessment and termination phases of the intervention appeared most vulnerable to inadequate strategy selection or competence in practice.

During the assessment phase this typically reflected the omission or delayed completion of one of the basic assessment tasks, rather than incompetent practice in a prescribed

area. The most commonly omitted task was the negotiation of a clear contract i.e. establishing the length of contact, implications of missed sessions etc. It was of note that more of the therapists trained later in the reviewed period of supervision did not complete this task on tape than in the Edinburgh group who were initially supervised. The majority of these therapists had had contact with their patients prior to starting to record their IPT casework for supervision, which none of the Edinburgh group had. It seems reasonable to assume from the discussions held during supervision that many of these therapists had outlined some of the practical details at this earlier stage, and then elected not to repeat this on tape. This is not the case for all therapists however as some found it difficult to manage the time limited contract and acknowledged their own confusion about session numbers, duration of treatment and missed sessions during supervision, reflecting the inadequacy of the contract which had been established with their clients. This may be an illustration of one of the difficulties of moving to a more clearly contracted model of work for therapists who are used to a more fluid arrangement.

The second most commonly missed task was the explanation of IPT and depression. For three therapists, who failed to address this task with either of the cases supervised, this appeared to reflect a theoretical conflict, in which the medical model of depression was unacceptable and not employed. Another therapist failed to meet almost all of the central assessment tasks, reflecting a significant difficulty in conducting the early part of the intervention in a manner consistent with manual guidelines. This may indicate a difficulty for some therapists in formulating patients' experience of depression in terms of the IPT model, and trying to implement IPT practice while continuing to think along more

familiar lines. This possibility is supported by the fact that all therapists who failed to explain the IPT model and depression reported that they preferred another model of treatment for depression. In all cases this included a more psychological rather than medical conceptualisation of depression. On four occasions a formulation was not negotiated during the first four sessions. While an area of interpersonal difficulty did emerge as the theme of the intervention in each of these cases, it was not decided through discussion of the focus options with the patient, potentially leaving the goals and expectations of the treatment unclear. This did not appear to be explained by the complexity of the presenting problems or difficulty in identifying an area of interpersonal difficulty. Each of the therapists who delayed or omitted to clarify the focus also omitted to set a clear general contract, suggesting a general difficulty in structuring and narrowing the remit of the episode of care being undertaken. Even with these deviations in practice, adherence and competence were not a significant problem for most therapists during the early sessions with approximately 80% of casework covering the majority of central tasks in a competent manner.

The issues related to termination received more variable attention, with 46% of cases being concluded without explicit discussion of the ending or its anticipated impact, or with only cursory reference to these issues. This appeared to be an area which generated some discomfort for a number of therapists, with some having to be explicitly directed to these tasks during supervision on a number of occasions before discussion was initiated with the patient. A number of therapists noted that prolonged and detailed attention to the conclusion of the therapy relationship deviated from their normal practice and was

consequently avoided, in what appeared to be a reflection of the termination reactions experienced by some of the therapists.

Three themes appeared to reflect the difficulties that arose around termination. The first was minimal symptomatic change. In some cases in which the patients reported little change in their experience of depressive symptoms the therapists appeared to avoid discussion of the unsatisfactory outcome, and continued to focus on what could be achieved during the final sessions. This resulted in a poor conclusion to therapy, without evaluation of the experience and may have been in danger of leaving patients feeling that they had failed therapy rather than therapy having failed to be of benefit to them. It is perhaps not coincidental that a number of these cases were concluded abruptly and prematurely, either with or without negotiation, suggesting a difficulty tolerating the end of the contact. This may also have denied therapists who were procrastinating, but not absolutely avoiding these tasks, the opportunity to address termination issues directly. These therapists were also more vulnerable to expressing feelings of having failed in providing an effective intervention with the new model of practice, and may have been trying to protect themselves as well as their patients from this conclusion. This may be a particular issue when the therapist is conducting a therapy for the first time, and does not have personal experience of value or efficacy of the new model.

A second factor which appeared relevant in a number of these cases was the noted deviation from normal practice in discussing the end of the therapy relationship. A number of the therapists were used to working in psychiatric settings with a fast turnover of patients, who were often not seen for psychotherapeutic work. In typical practice for

these therapists decisions about the end of contact would be made more quickly and on the basis of most recent presentation. This shift in attention to include discussion and review of the completed work and therapy relationship appeared difficult for some regardless of the outcome of treatment.

A third issue was the overlap between those therapists who had difficulty establishing a contract in the early sessions and those who had difficulty addressing the conclusion of the contact at the end of treatment adequately and competently. It appeared that the responsibility for managing the structure of the intervention and directing the focus during the distinct phases of treatment was generally difficult for these therapists, with the result that tasks were bypassed or given only passing reference.

In contrast the middle sessions and focus directed work was completed in a manner consistent with manual recommendations by the majority of therapists, in the majority of cases. The themes addressed were shown to vary consistently with the interpersonal foci selected, and the dual goals of symptom reduction and resolution of interpersonal difficulties were addressed and linked in the majority of sessions by the majority of therapists. It appears then that most therapists were able to conduct the focus work in a manner consistent with manual recommendations, but adapting to the different demands of the three phases of treatment proved problematic for many. It would appear then that the majority of therapists achieved adequate adherence and competence with regard to the formulation based work but did not consistently maintain this standard across the different phases.

This program of training and supervision followed the procedures employed during the NIMH TDCRP (Rounsaville et al, 1986), and replicated their finding that therapists could learn to perform IPT to an acceptable level under supervision, following review of the treatment manual and completion of a one week didactic training course. The therapists in the present study performed at a mean competence level consistent with highly selected and experienced psychodynamic psychotherapists, and above that reported for the less experienced comparison group. In comparing the range of initial and termination strategies undertaken, the subjects in the present study made use of fewer prescribed strategies than those reported by Rounsaville et al (1986) i.e. 60%: 68% and 57%: 70%, suggesting a lower standard of adherence in the present sample. This may reflect the greater diversity of general training and theoretical backgrounds represented in the present sample, and the greater accommodation necessary for therapists who have not been selected at least partly because their routine practice compliments the training being completed. This may also reflect the greater number of training cases conducted by most of the therapists in Rounsaville et al (1986), compared with those in the present study. Therapists in the 1986 study completed up to four training cases, while only two of the therapists in the present sampled conducted this number of supervised cases. The majority only had the opportunity to conduct a first and second case, and therefore did not have the same opportunity for improvement through experience as may have been illustrated in the earlier sample. Thus therapists in the present study may be unduly criticised in this comparison.

While the mean ratings for all therapists reflected an acceptable standard of practice, five of the therapists did not achieve a sufficiently high and consistent standard of

practice to meet requirements. The quality of some of the individual interventions were rated as adequate for some of these therapists, but the IPT specific interventions were used with low frequency, and in several examples missed out entirely. Consequently the treatments could not be regarded as characteristic of the model of practice described in the IPT manual. Three of these therapists also made inconsistent use of the supervision available, resulting in prolonged gaps between meetings. This may have reflected a reluctance to adapt practice from another well practised model, and seems likely to have resulted in missed opportunities to redirect patterns of intervention which moved away from the standard IPT treatment. Unfortunately these therapists did not take on second IPT training cases and so it was not possible to establish whether subsequent supervised casework would have resulted in an increase in the application of the prescribed IPT techniques.

It is however important to provide a context for this finding and it should be noted that those therapists who did not meet the competence standard all fell just below the acceptable range, which was set at a mean competence rating of 3.5 and above. This study sets a reasonably high standard at the cut off between the higher and lower end of the seven point evaluation scale. According to Rounsaville et al (1986) ratings below this mid point cut off would still be regarded as performing at a "good" level of competence, but in the absence of normative scores a conservative standard was adopted. This judgement illustrates the vulnerability inherent in employing scales without a manual and normative scores to facilitate evaluation of the resulting scores. The lack of attention given to promoting the use of standardised scales for IPT evaluation in a population beyond the authors is a major drawback, bringing into question continuity of

anchor points employed for ratings and comparative nature of results generated by different groups. Anchor points were provided for each of the raters who contributed to this study in order to achieve some standardisation, but the discontinuity between the conclusion reported in Rounsaville et al (1986) and the present study suggest that comparisons between the two samples may be unsafe.

10.2 INTERRATER RELIABILITY

Given that most of the therapists in the study were trained and supervised by the same person, it was important to determine whether the same ratings would have been generated by someone independent of the training and supervision relationship. In addition the fact that reports on the rating forms which were employed had previously been produced by clinicians and researchers closely involved in their development, left the question of generalisability unanswered.

The general correspondence between supervisor and external raters, which was demonstrated both for ratings of adherence and competence was adequate in the present study. A clear majority of process, strategy and focus specific items were consistently rated to have been present or absent by the rating dyads, and this was reflected in acceptable Kappa ratings for all of the scales. In addition the majority of items were rated within one point of the original rating when competence scores were compared. This generated a high level of agreement across raters on the acceptability of performance across therapists. However when correlations between ratings were calculated it became evident that the trend of scoring on individual scales was not consistent and the second raters were as likely to rate performance as better or worse

than the judgement made by the first rater. This generated an inconsistent pattern, with approximately half of the ratings being made in each direction on the PRF and the focus forms. This reflects one of the noted vulnerabilities of the forms employed in this study, which is the lack of item definition. Raters appear able to make consistent general judgements about the quality of the work reviewed but the specific quality is either judged differently or is rated according to a different interpretation of the scale. Without clear manual guidelines it is difficult to substantiate either conclusion robustly. In order to understand the apparent discrepancy between the general correspondence percentages and the correlations the raw data was examined in detail. This revealed a highly consistent pattern of scoring for each subject, with all therapists consistently being scored slightly higher or slightly lower across the scales by the second rater, which may explain the significant correlation on the composite score but not the individual PRF or focus forms. Eighty percent of these ratings were within a one point range. Although not entirely broken down along individual rater lines it did appear that there was a trend for the individual second raters to rate with slightly higher scores in one instance and slightly lower scores in the other, across their sample of tapes. This suggests that the lack of clear anchoring definition for the items rated was influential in the rating patterns which emerged, demanding that idiosyncratic definitions be created when universal guidance was not available. The small sample size and skewed distribution must also be considered as potentially influential factors.

Each tape was second rated by only one of the external raters to increase the sample of tapes which could be double rated and because Foley et al (1987) reported that double rating was sufficient to achieve an adequate measure of reliability. Despite these efforts

to extend the external rating as far as possible it must be acknowledged that this requires further validation. Not all therapists or patients in the sample were reviewed by a rater who was independent of the supervision relationship, and the presented results can therefore only be taken as an indication of the reliability of scoring for a sample of the population under study. It was also evident that the individual raters had a tendency to rate certain items with different frequency e.g. "use of significant other" was used regularly by one rater but not at all by the other two. It became evident in discussion among the raters, following completion of the scoring exercise, that not all items were readily or consistently understood, and the lack of manual guidelines leaves rating vulnerable to misinterpretation or underrating.

Another difficulty which was noted, was how to use the lower end of the rating scale appropriately. This was not a frequent difficulty as most of the casework was rated to be acceptable, but a number of the discrepancies which were evident in the ratings related to uncertainty about rating specific interventions as present but inadequate or absent. Once again the lack of consistent guidance became evident in these instances. The anchor points employed by the original supervisor were provided for the external raters to facilitate agreement but it cannot be assumed that this system of interpretation reflects the previously published data on these evaluation forms, and previous discussion of Rounsaville et al's (1986) conclusions about good standards of practice would suggest some inconsistency.

Limiting the second rating to a single tape may also have weakened the reliability exercise. This meant that only a proportion of strategies were reviewed, and although

there was good agreement about what was evident in the session and what was not, the difficulties inherent in rating a wider range of interventions may not have been fully tested. Each of the focus areas were included in the second rating sample and this matched the proportion of cases in each focus in the whole sample. Similarly the majority of strategies were evident at some point in all the casework rated but the difficulty in completing more sophisticated statistical analysis indicates that the range may have been insufficient to robustly test the question of reliability.

It was encouraging to note that review of the notes made by the external raters during the rating exercise corresponded well with the opinion of the supervisor, with similar themes being noted and common lines of criticism of practice. This allowed the supervisor to accurately identify the therapists and patients discussed when the tapes were still anonymous. It would appear then that adequate reliability was established to validate the conclusions of this study, but much more work is required in this area to adequately test the appropriateness of the current forms as the routine basis for IPT supervision.

10.3 COMPETENCE AND OUTCOME

This study failed to demonstrate a relationship between the competence of practice and clinical outcome. This connection has found varying levels of support previously in the IPT literature (Rounsaville et al, 1981; O'Malley, 1988, Frank et al 1991), and the definition employed on each side of the equation has been shown to be very influential. Rounsaville et al (1981) also failed to reveal a relationship between process and outcome. However, as with the competence rating for the early middle

session in the present study, performance appeared to be significantly related to outcome when univariate statistics were employed but was subsequently shown to be non significant when initial patient characteristics were controlled. It was of interest that a significant relationship was suggested on the basis of an individual session but not the overall rating. This may suggest that the composite score collapsed too much information to be sensitive to the subtle relationship which may be played out between dimensions of practice and outcome. This hypothesis would be consistent with O'Malley et al's (1988) study, which examined competence on the basis of a single session and found a significant correlation with outcome, although measurement differences also have to be taken into account. This also demonstrates that patient characteristics such as initial symptom severity and even broad indicators of engagement in therapy, such as number of sessions attended held a closer relation to the outcome of treatment than the specific practice of the treatment itself in the present sample.

O'Malley et al (1988) did find a relationship between process and outcome, but it was of interest to note that therapist performance made the greatest contribution to patient rated change rather than symptom based outcome ratings. This measure would appear most likely to reflect some judgement on the therapy relationship and satisfaction with this, which may hold an uncertain relation to the degree of symptom change. Given the specific finding on the apathy rating of the HRSD in relation to competence, it may be that competent practice facilitates a more enthused or interested attitude in patients, towards themselves and their circumstances which is not entirely dependent on global symptom change. The strategy and symptom specific ratings employed in the present

study would be unlikely to capture this motivational dimension and so may have failed to attend to the mechanisms through which practice and outcome may be linked.

The BDI-ii has not previously been employed to measure outcome in relation to practice competence in IPT, with more functional definitions of outcome and clinician rated symptoms scored being employed in its place, and the BDI-ii may be an insensitive measure to use in exploring this relationship. The BDI-ii conceptualises depression in a more primarily cognitive manner and may therefore offer a less direct measure of the nature of the change achieved over the course of an IPT treatment. Many of the sample supervised were not familiar with the Hamilton Rating Scale for Depression, which would have been an alternative depression rating to employ, and it was decided not to employ this as an outcome measure as ratings may not have been consistent across therapists.

It would seem unlikely that the general pattern of symptom improvement reported for the sample as a whole masked the relationship between competence and outcome, as there was reasonable variance across patients. An alternative interpretation of this finding however is that given general standard of practice was high for the majority of therapists, it may be the lack of variance in performance which prevent significant findings emerging from the data, rather than a lack of relationship between the specified variables. The composite score which was employed does appear vulnerable to missing the detail which may be of value in charting the associations between process and outcome. Numerous different patterns of practice can be represented within a competent score and general competence feels inadequate as a concept to capture the diversity this

represents. However on the basis of these analyses it would appear that the competence of clinical practice is not as significant with regard to outcome as it might intuitively appear, and is superseded by patient and engagement factors.

10.4 THERAPEUTIC ALLEGIANCE

Rounsaville et al's (1986) suggested that superior performance during training cases may reflect similarities between the new skills being taught and the existing experience of trainees, with psychodynamically trained therapists best able to adapt to IPT requirements. However the therapists in the present study came from a variety of therapeutic origins, with only one being psychodynamically trained, and professed a preference for a range of therapeutic styles, yet they were rated as performing to an equally high standard on their initial training cases as the psychodynamic therapists. The findings of this study suggest that experience of and allegiance to a variety of models of treatment for depression may at best assist therapists in learning to use IPT, and at least present no insurmountable obstacles to doing so.

This may also reflect the timing of the ratings as the therapists in this sample were being introduced to this model and were undertaking their initial casework. It is therefore perhaps surprising that so many therapists were sufficiently impressed and convinced by the model that they reported it to be their preferred model at this early stage. This may reflect early enthusiasm for alternative interventions immediately following training. In previous reports on therapeutic allegiance the therapists under study had chosen to conduct research evaluating specific models of therapy, while the therapists in this sample had chosen only to complete training in the model. It is also important to note

that the majority of therapists did include IPT in their list of preferred models, and so attributing a stronger preference to those who placed it first on the list may have created an arbitrary division in a homogenous group. It may also be important to note that the unqualified therapists in the current sample had a mean of 2.8 years experience, much of which would not have involved client work, and had not completed any formal psychotherapy training in any model. Thus the preference expressed by almost 40% of the sample may have been theoretical rather than experience based, and may therefore have generated an easier obstacle to overcome.

Chevron et al (1983) in their paper on selecting psychotherapists to perform IPT suggest that therapists should be, 1) fully qualified psychiatrists or psychologists, 2) have had a minimum of two years of experience following completion of training, 3) have received training in a psychodynamically oriented framework and 4) have treated a minimum of 10 depressed patients in psychotherapy. The present study suggests therapists experienced in treating depression from either a CBT or psychodynamic perspective are equally able to learn and implement the procedures specified in the IPT manual. This is encouraging giving the growing interest in and demand for training in this model from a wide selection of caring professionals.

It may be that therapists from different backgrounds experience different challenges in adapting their practice. During supervision it was of note that therapists with more experience in CBT found no difficulty in working within a time limited and focused intervention, as this would presumably be part of their routine practice. Greater difficulty emerged however in maintaining a consistent interpersonal focus, rather than resorting

to an exploration of the biases and maladaptive thinking processes which lay behind interpersonal exchanges. Although CBT does not avoid reflecting on affect, the continual focus and processing of affect demanded in IPT often proved difficult for these therapists to implement. More psychodynamically oriented therapists on the other hand worked in this area with ease, but were concerned that the intervention was rushed and overly directive, and did not allow adequate examination of the historical origins of presenting difficulties.

During supervision one of the comments frequently made by therapists was that they found themselves initially restricted by the demands of the training exercise. They were concerned when the protocol would not allow them to implement the range of interventions which they might normally have relied upon. Therapists expressed concern that in trying to stick to the prescribed techniques they may have disadvantaged their clients by delivering a clunky and mechanistic intervention which was not adequately responsive to their needs. These training concerns reflect very closely the criticisms presented by Persons (1991), and Henry et al's (1990) description of therapists' experience of training. Crits-Christoph (1998) reported that in IPT the treatment focus was less flexibly implemented than in CBT. CBT was shown to allow greater responsiveness to the patient's beliefs about the important factors in formulating their experience of depression. Whereas IPT works from a clearly established assumption that interpersonal factors play a significant role in the presentation of a depression and consequently are pursued independently of the patient's belief in this perspective. It was of interest in the present study that such comments and concerns reduced over the course of the interventions, apparently as therapists' confidence in the

range of techniques available to them, and presumably their ability to implement them competently, grew. Only two of the twenty eight cases were judged by the therapists to have been formulated around an inappropriate focus at the end of treatment.

10.5 CLINICAL EXPERIENCE

Several definitions of clinical experience were employed to explore the impact of this variable on therapist adherence and competence ratings and clinical outcome. Few were found to correlate. This is in direct contrast to Rounsaville et al's (1986) finding on the superiority of experienced therapists in initial training cases, and Chevron et al's (1983) proposal that more experienced therapists had greater potential as IPT therapists. However it should be noted that the previous reports of experience were based on highly selected samples and involved numerous assumptions beyond that expressed in terms of duration of clinical practice. Therapists in the early IPT studies were selected on the basis of previous psychodynamic training, favourable attitude to IPT and rejection of attitudes and practices inconsistent with the IPT model. No such assumptions were met in the present sample and previous clinical practice reflected a variety of models of therapy and intervention, and for many these would not be primarily characterised as psychotherapy. Parloff et al (1978) suggested that the lack of experience effect in some studies may be related to an inadequate standard by which to categorise therapists as experienced. Thus the independence of adherence, competence and experience may reflect the insensitivity of the basic definitions of experience employed, particularly when some categories were dichotomised to provide sufficient numbers for comparison, and the more heterogeneous character of the sample. The findings of this study do not specifically support the role of duration therapists' general experience or number of

depressed patients treated in additionally facilitating their performance of a new model of therapy. In practice these dimensions did not generate uniform groups and did not highlight the aspects of experience which might be thought to be of value in conducting psychotherapy.

Auernbach & Johnson (1977) discuss the difficulties inherent in trying to conceptualise a continuous variable such as experience, which may be defined in a variety of ways, particularly when research typically demands that it be categorised i.e. experienced/inexperienced, and associated with skills with which it may or may not bear a linear relationship. Chevron & Rounsaville (1983) proposed criteria by which therapists may be selected as good candidates for IPT training, and they found that older, more experienced therapists (defined by years of practice) were rated as more empathic and with greater potential as IPT therapists. If the therapists in the current study were compared with the data presented by Chevron & Rounsaville (1983) they would in fact be classed as unacceptable for training, based on their mean age and level of clinical experience as mean ratings on both dimensions fell within the range reported for therapists considered to be poor candidates for IPT training. The definition of experience employed is therefore not borne out as a valid predictor in light of evaluated casework with a less selected sample of therapists. This is encouraging in that it suggests that the more diverse population now applying to and completing IPT training cannot be assumed to perform poorly because they do not meet rigorous research inclusion criteria, but it also makes it difficult to reliably determine who may be able to make use of the limited training resources which are available.

Of more value in this regard was the finding that all of the therapists who failed to meet the overall standard of competence and adherence had a psychiatric training background, and only one of them routinely conducted psychotherapy with the majority of patients seen in normal practice. None of the psychology trained therapists failed to meet criteria. This demonstrates that competence ratings were significantly related to the type of professional training therapists had. This is inconsistent with Chevron et al's (1983) finding that professional degree was unrelated to ratings on IPT potential in therapists applying for training, but once again it was assumed that therapists of all professional disciplines would have completed psychodynamic training and two years practice prior to applying, which may be seen to outweigh original training.

The current study largely failed to replicate Blatt et al's (1996) finding that more effective therapists had a more psychological rather than a biological orientation to their standard treatment process. Some modest evidence emerged to support the greater efficacy of unqualified psychology graduates over psychiatry trained therapists, but this was based on small numbers and was not supported by the general efficacy of qualified psychologists, although the trend in the mean data supported this proposal. In contrast this study did provide evidence of this division in the competence with which treatment was delivered. It should be noted however that Blatt et al's (1996) classifications were based on more detailed information elicited from therapists on their routine practice than was available in the present study. Blatt et al's (1996) sample may also have provided more distinct groups, as the present sample were all likely to work with patients using combined pharmacological and psychological interventions, although the respective biases may have differed. Although not specifically addressed in this study it would be

of interest to know if this training background split also reflected a difference between the therapists in psychotherapy experience. This would be the typical treatment provided by psychologists, while many of the medical and nursing, psychiatry trained participants worked in the general psychiatry setting and conducted psychotherapy only in special interest sessions which accounted for a small proportion of their clinical time or in conjunction with a high proportion of medically oriented treatment.

10.6 PATIENT FACTORS

The influence exerted by patients on practice and outcome was explored in terms of initial symptoms severity, early symptom response to treatment and in therapy problems. Previous reports on IPT (Elkin et al,1985) found that this model worked well as a treatment for both moderately and severely depressed patients, and the results of the current study support this, with no significant differences emerging between the two groups on either end point ratings or extent of clinical change. Greater early change in symptoms was also related to better outcome, but was unrelated to overall competence. This suggests that clinical change is achieved by a route unrelated to the competence of the therapy delivered, although this may only hold true if the therapy is above a certain threshold as reported in the current study.

The rating made of problems with conducting therapy proved important, demonstrating an inverse relationship with both measures of clinical outcome and early competence in delivering therapy. These results are consistent with Foley et al's (1987) findings on patient difficulty. Although the rating methods are not the same they may tap into a similar dimension. Foley et al (1987) defined difficulty in terms of help rejecting

behaviour and the present study recording lateness and missed appointments and an unwillingness to engage in the contracted work in IPT e.g. to maintain attention to a focus area and to discuss interpersonal relationships. In so far as these examples reflect on the collaborative nature of the therapy exchange these findings correspond with the association between therapeutic alliance and outcome. The global rating employed in the present study could not be regarded as a direct measure of alliance, but does record the frequency with which alliance threatening behaviours were performed, and therefore may trace the potential for ruptures to the alliance across the treatment interval. This would provide a means to explain the correspondence between the results of the present study and Krupnick et al (1994) who found that in IPT specifically the strength of the alliance significantly distinguished the most and least improved cases, and that the most influential contribution to the alliance was that of the patient.

It is of course important to consider the possibility that the findings on problems in therapy are biased by observation of clinical change i.e. is therapy which is clearly effecting a change rated as less problematic? While this is a possible confounding factor, it is important to note that the majority of the ratings were simple recordings of events e.g. missed appointments, late attendance, reports of suicidal intent or acts, early termination of therapy, and are not open to interpretation. In addition the symptom reports by patients are beyond the control of the supervisor and therefore cannot be biased into higher ratings because of an awareness of therapy events as might be the case with clinician administered scales like the HRSD. As with all ratings reported in this study the problems apparent in each session were recorded after each tape had been reviewed and prior to the supervision meeting. Therefore all ratings were made

independent of knowledge of final outcome, although it must be noted that the clinical evaluation of the patient's wellbeing made when the tape was reviewed may exert an influence over the more interpretative items.

10.7 COMPETENCE RATINGS

Waltz et al (1993) set out a number of recommendations on practice when evaluating therapist adherence and competence in implementing a psychotherapy protocol. It is possible to evaluate to what extent this exercise implemented these recommendations:

1) Define competence in relation to the treatment manual.

This was done by using rating scales which had been developed with specific reference to the IPT manual, and which used the descriptions of strategies employed by the manual. This maximised the chances of therapists and supervisor sharing a common understanding of the strategies being evaluated. These rating forms had previously been employed in the largest scale evaluation of IPT - the NIMH TDCRP, and the independent raters on that study had been chosen from the originators and authors of the IPT manual and judgements were substantiated by corroboration with independent evaluators (Rounsaville et al, 1986).

2) Tailor manipulation checks to the questions being asked.

The specific objective of this exercise was to establish whether or not therapists could be trained to consistently and competently implement the IPT strategies, following training and supervised casework. To this end the checks employed were directly tailored to meet this need, rating both the use of the individual strategies and the quality

with which they were implemented. This exercise focused only on IPT training cases and so there was no need to specify alternative styles of intervention, as would have been possible with the Collaborative Study Psychotherapy Rating Scale (CSPRS; Hollon et al, 1988), and would have been important had the objective been to establish how consistently two forms of therapy could be distinguished. Although only one form of therapy was under study, note was taken of the frequency with which strategies more characteristic of an alternative form of therapy e.g. psychoanalytic interpretations, were employed, as DeRubeis et al (1982) noted that it is important both to establish what should be there is, and what should not, is not.

3) Include unique and essential behaviours, essential but not unique behaviours, compatible but not necessary or unique and prohibited behaviours.

This range is covered in the ratings, although the focus is more specifically on the essential and compatible ratings with only a single reference to prohibited strategies. Ratings were made of both the focus specific interventions e.g. linking the onset of symptoms to the negotiated interpersonal focus, and those which would characterise the general process of psychotherapy but could not be considered unique to IPT e.g. exploratory techniques, discussion of affect. An alternative could have been employed by using the CSPRS (Hollon, 1988) which details IPT strategies, CBT strategies and medication review, but the size of that scale was prohibitive and was not as immediately accessible for supervision purposes, which was the original purpose of the ratings.

4) Assess therapist competence.

The present study rated not just the presence or absence of IPT strategies i.e. adherence, but also made explicit ratings of the quality of the interventions. In order to

make an informed evaluation of competence, Waltz et al (1993) recommends that ratings should take into account what session is being assessed and what has happened in previous sessions, issues of patient resistance and the focus of the intervention. Given that the ratings were made during the course of approximately weekly supervision throughout treatment all of the information suggested was available when ratings were made. In this way attention was not given simply to the presence or absence of specific techniques, but also to the appropriateness of the intervention (Hill et al, 1981, Tracey et al, 1981), and the skill with which it was implemented.

10.8 LIMITATIONS OF THE PRESENT STUDY

The relative inexperience of the supervisor in this study meant the skill with which this role was fulfilled was developing through the course of the exercise. This was particularly important given the lack of manual guidelines for the rating measures which were employed. Guidelines provided in the therapy manual were employed whenever possible, but these are limited and left a range of items open to more unstandardised observation. Observation of the patterns of scores across the duration of the study do raise the possibility that there was some drift in ratings over the three years during which they were produced. It appears that there were a higher proportion of strategies being rated as present but below competence standards rather than absent in later cases. This has been noted in discussion with the other raters as a matter of uncertainty in using the IPT scales.

Although the rating system in the present study is vulnerable to criticism it does reflect the current practice of IPT supervisors who have taken up the supervision forms as a guide to their evaluation. Insofar as it provides a measure of the comparative nature of the evaluations made by current supervisors it is hoped that the study provides valuable information. This however is acknowledged as only a preliminary step in producing a standardised system of therapy evaluation with normative scores to gauge the performance level of both new and experienced IPT therapists.

A further point of note is that there were no comparable ratings on practice prior to the training and supervision being completed. It is therefore not possible to firmly conclude that the IPT skills demonstrated are consequent to the training and supervision. This impression relies in addition on the enthusiastic, and at times exasperated, comments frequently made by therapists during supervision that their IPT practice focused more consistently and in greater depth on the interpersonal context and affect experience of their patients.

Another potentially influential factor was the context of the symptom rating. All ratings were completed at the request of the therapist during or immediately prior to the therapy session. This is a potentially emotionally laden situation and may be vulnerable to communicating more than the average experience of symptoms over the last week. As scores were also collected directly by the therapist, rather than through the anonymity provided by an independent rater, they may have been at least in part, illustrative of feelings about the therapeutic relationship. This may in part explain the significant relationship demonstrated between perception of difficulty in conducting therapy and

outcome. The final score may be particularly vulnerable to this effect given that it was completed during the final therapy session. Such scores may be influenced by the impact of issues related to termination, which may be difficult to disentangle from the general symptom level in the preceding week. These scores may not be different in cases in which termination was particularly difficult. Ideally scores would have been taken again at one month post therapy. This would have allowed some distance from the conclusion of therapy and may have given a clearer indication of the remaining symptom difficulty.

As previously noted the lack of an evaluation manual to guide the use of the IPT forms is a fundamental difficulty faced by the present study. Although acceptable reliability was demonstrated, the concordance between the standards and definitions applied by the current raters and the originators of the forms remains uncertain. This study can provide only provisional evidence of the suitability of the IPT forms as a basis for supervision and a more substantial sample rated over a greater range of therapy sessions would be necessary to produce a more robust conclusion.

10.9 CONCLUSION

This study demonstrated that practising therapists, with a range of experience and theoretical influences, were able to practice the procedures outlined in the Interpersonal Psychotherapy manual (Klerman et al,1984), with a high level of competence. In addition current IPT supervisors demonstrated a acceptable level of agreement on the presence and absence of the prescribed techniques and reliably rated the majority of interventions

as adequate or inadequate. Adherence levels were good in the focus area sessions, but less satisfactory during the initial and final phases of treatment. Less experienced therapists were found to be as capable of meeting training requirements as more experienced therapists, and a significant level of symptomatic relief was reported by the participating patients. Initial symptom severity did not have a detrimental effect on treatment outcome, with patients rated as severely depressed on the BDI-ii at baseline achieving recovery or clinically significant reduction in symptoms as often as patients with a moderate depression. Therapists with a psychology based training achieved a higher standard of competence than therapists trained in a psychiatry model of medicine or nursing, but the two groups could not be distinguished in terms of clinical outcome for patients. Problems in conducting therapy, reflecting potential ruptures in the therapeutic alliance were significantly related to clinical outcome and early competence.

Appendix 1

Interpersonal Role Transition

Goals

1. Mourning and acceptance of the loss of the old role.
2. Help the patient to regard the new role as more positive.
3. Restore self-esteem by developing a sense of mastery regarding demands of new roles.

Strategies

1. Review depressive symptoms.
2. Relate depressive symptoms to difficulty coping with some recent life change.
3. Review positive and negative aspects of old and new roles.
4. Explore feelings about what is lost.
5. Explore feeling about the change itself.
6. Explore opportunities in the new role.
7. Realistically evaluate what is lost.
8. Encourage appropriate release of affect.
9. Encourage development of social support system and of new skills called for in new role.

Grief

Goals

1. Facilitate the mourning process.
2. Help the patient re-establish interest and relationships to substitute for what has been lost.

Strategies

1. Review depressive symptoms.
2. Relate symptoms onset to death of significant other.
3. Reconstruct the person's relationship with the deceased.
4. Describe the sequence and consequences of events just prior to, during and after the death.
5. Explore associated feelings (negative as well as positive).
6. Consider possible ways of becoming involved with others.

Appendix 1

Disputes

Goals

1. Identify the dispute.
2. Choose plan of action.
3. Modify expectations or faulty communication to bring about a satisfactory resolution.

Strategies

1. Review depressive symptoms.
2. Relate symptoms' onset to overt or covert dispute with significant other with whom the patient is currently involved.
3. Determine stage of the dispute:
 - a. renegotiation (calm down participants to facilitate resolution);
 - b. impasse (increase disharmony in order to reopen negotiations);
 - c. dissolution (assist mourning).
4. Understand how nonreciprocal role expectations relate to dispute:
 - a. what are the issues in the dispute?
 - b. what are the differences in expectations and values?
 - c. what are the options?
 - d. what is the likelihood of finding alternatives?
 - e. what resources are available to bring about change in the relationship?
5. Are there parallels in other relationships?
 - a. what is the patient gaining?
 - b. what unspoken assumptions lie behind the patient's behaviour?
6. How is the dispute perpetuated?

Deficits

Goals

1. Reduce the patient's social isolation.
2. Encourage formation of new relationships.

Strategies

1. Review depressive symptoms.
2. Relate depressive symptoms to problems of social isolation or unfulfillment.
3. Review past significant relationships including their negative and positive aspects.
4. Explore repetitive patterns in relationships.
5. Discuss patient's positive and negative feelings about therapist and seek parallels in other relationships.

Appendix 1

Termination Strategies

1. Explicit discussion of the end of treatment
2. Elicit/discuss patient's reaction to termination
3. Acknowledgement of the end of treatment as a time of potential grieving
4. Help patient move toward recognition of his/her independent competence
5. Review the course of treatment and progress with the patient
6. Patient given opportunity to evaluate the treatment and assess future needs
7. Assess with patient early warning signals, and discuss procedures for re-entry into treatment if necessary.

Appendix 2

INITIAL SESSION(S) CHECKLIST AND RATING FORM

Patient:

Therapist:

Therapy Date:

Rater:

Consultation Date:

Therapy Session:

TASKS

QUALITY

	Yes	No	Rec	Excellent				Poor			
Inquire re: chief complaint and depressive symptoms	1	2	3	1	2	3	4	5	6	7	
History of current depressive episode	1	2	3	1	2	3	4	5	6	7	
Inquire re: previous history of depressive episodes and treatment	1	2	3	1	2	3	4	5	6	7	
Brief Social History	1	2	3	1	2	3	4	5	6	7	
Give the syndrome a name	1	2	3	1	2	3	4	5	6	7	
Explanation of IPT & Depression	1	2	3	1	2	3	4	5	6	7	
Give the patient the "sick role"	1	2	3	1	2	3	4	5	6	7	
Reassurance re: positive prognosis	1	2	3	1	2	3	4	5	6	7	
Evaluate need for medication	1	2	3	1	2	3	4	5	6	7	
Inquire re: patient's expectations about psychotherapy	1	2	3	1	2	3	4	5	6	7	
Translation of chief complaint (depressive symptoms) into interpersonal context	1	2	3	1	2	3	4	5	6	7	
Interpersonal Inventory	1	2	3	1	2	3	4	5	6	7	
Feedback IPT Formulation/ Identify Focus	1	2	3	1	2	3	4	5	6	7	
Explanation of IPT techniques	1	2	3	1	2	3	4	5	6	7	
Contract setting (admin. details, length, frequency and duration of sessions and treatment, appointment times)	1	2	3	1	2	3	4	5	6	7	
Contract setting re: treatment goals	1	2	3	1	2	3	4	5	6	7	
Exploration of therapist and patient tasks working towards treatment goals	1	2	3	1	2	3	4	5	6	7	

Therapist Behaviour

Quality

	Excellent					Poor	
Appropriate Activity Level	1	2	3	4	5	6	7
Appropriate Degree of Supportiveness	1	2	3	4	5	6	7
Focus on current interpersonal functioning	1	2	3	4	5	6	7

Comments

Appendix 2

INTERPERSONAL DEFICITS RATING FORM

Patient:

Therapist:

Therapy Date:

Rater:

Consultation Date:

Therapy Session:

Goal Directed Activity

Quality

	Yes	No	Rec	Excellent				Poor			
Review depressive symptoms	1	2	3	1	2	3	4	5	6	7	
Relate depressive symptoms to problems of social isolation, or social unfulfillment	1	2	3	1	2	3	4	5	6	7	
Review past significant relationships including negative and positive aspects	1	2	3	1	2	3	4	5	6	7	
Explore repetitive dysfunctional patterns of behaviour and/or expectations in relationships	1	2	3	1	2	3	4	5	6	7	
Discuss patient's positive and negative feelings about therapist and explore parallels in other relationships	1	2	3	1	2	3	4	5	6	7	

Comments

INTERPERSONAL ROLE DISPUTES RATING FORM

Patient:

Therapist:

Therapy Date:

Rater:

Consultation Date:

Therapy Session:

Goal Directed Activity

Quality

	Yes	No	Rec	Excellent				Poor			
Review depressive symptoms	1	2	3	1	2	3	4	5	6	7	
Relate symptom onset to overt or covert dispute with significant other with whom the patient is currently involved	1	2	3	1	2	3	4	5	6	7	
Determine the stage of the dispute	1	2	3	1	2	3	4	5	6	7	
Identify issues in the dispute	1	2	3	1	2	3	4	5	6	7	
Explanation of how non-reciprocal role expectations relate to the dispute	1	2	3	1	2	3	4	5	6	7	
Exploration and discussion of differences in expectations and values	1	2	3	1	2	3	4	5	6	7	
Exploration of parallels in other relationships	1	2	3	1	2	3	4	5	6	7	
Exploration and discussion of options available to the patient	1	2	3	1	2	3	4	5	6	7	
Discussion of communication patterns (structural, emotional, expectational and wish aspects)	1	2	3	1	2	3	4	5	6	7	
Exploration and discussion of how dispute is perpetuated	1	2	3	1	2	3	4	5	6	7	

Comments

Appendix 2

GRIEF AND LOSS RATING FORM

Patient:

Therapist:

Therapy Date:

Rater:

Consultation Date:

Therapy Session:

Goal Directed Activity

Quality

	Yes	No	Rec	Excellent				Poor			
Review depressive symptoms	1	2	3	1	2	3	4	5	6	7	
Relate symptom onset to death of significant other	1	2	3	1	2	3	4	5	6	7	
Reconstruct the patients relationship with the deceased	1	2	3	1	2	3	4	5	6	7	
Describe the sequence and consequences of events just prior to, during and after the death	1	2	3	1	2	3	4	5	6	7	
Evaluate availability and use of social supports around mourning	1	2	3	1	2	3	4	5	6	7	
Explore associated feelings (negative and positive)	1	2	3	1	2	3	4	5	6	7	
Consider alternative ways of becoming involved with others	1	2	3	1	2	3	4	5	6	7	

Comments

ROLE TRANSITIONS RATING FORM

Patient:

Therapist:

Therapy Date:

Rater:

Consultation Date:

Therapy Session:

Goal Directed Activity

Quality

	Yes			No	Rec.	Excellent				Poor		
Review depressive symptoms	1	2	3			1	2	3	4	5	6	7
Relate depressive symptoms to difficulty in coping with some recent life change	1	2	3			1	2	3	4	5	6	7
Review positive and negative aspects of old role and possible new ones	1	2	3			1	2	3	4	5	6	7
Explore feelings about what is lost	1	2	3			1	2	3	4	5	6	7
Explore feelings about the change itself	1	2	3			1	2	3	4	5	6	7
Explore opportunities in new role	1	2	3			1	2	3	4	5	6	7
Realistic evaluation of what is lost	1	2	3			1	2	3	4	5	6	7
Encourage appropriate release of affect	1	2	3			1	2	3	4	5	6	7
Encourage development of social support system and new skills called for in new role	1	2	3			1	2	3	4	5	6	7

Comments

Appendix 2

PROCESS RATING FORM

Patient:

Therapist:

Therapy Date:

Rater:

Consultation Date:

Therapy Session:

PROCESS

QUALITY

	Yes No Rec			N/AExcellent								Poor	
Exploratory Techniques: Supportive Acknowledgement; Extension of topic; Non-Directive Exploration	1	2	3	0	1	2	3	4	5	6	7		
Administrative Details	1	2	3	0	1	2	3	4	5	6	7		
Encourage expression of Affect; Inquiry into sensitive areas; Acceptance/Acknowledgement of Affect; Inquiry into feeling associated with content	1	2	3	0	1	2	3	4	5	6	7		
Clarification/Confrontation; Restructuring, Rephrasing, Feedback; Development of Interpersonal Awareness; Interpretation	1	2	3	0	1	2	3	4	5	6	7		
Communication Analysis	1	2	3	0	1	2	3	4	5	6	7		
Use of the Therapeutic Relationship	1	2	3	0	1	2	3	4	5	6	7		
Directive Techniques; Advice Giving; Limit setting; Education; Modelling; Direct Help.	1	2	3	0	1	2	3	4	5	6	7		
Decision Analysis	1	2	3	0	1	2	3	4	5	6	7		
Use of Significant Other	1	2	3	0	1	2	3	4	5	6	7		
Other	1		3	0	1	2	3	4	5	6	7		
Non-IPT Techniques e.g. Behaviouristic, Overly Psychoanalytic	1	2	3	0	1	2	3	4	5	6	7		

Client Therapist

Problems With Doing Therapy

	Yes	No	Yes	No
Lateness	1	2	1	2
Missed Appointments	1	2	1	2
Changed Topic/Tangential	1	2	1	2
Direct Unco-operativeness	1	2	1	2
Excessive Dependency/Demands	1	2	1	2
Suicide Threats	1	2	1	2
Early Termination Threats	1	2	1	2
Distorted View of Therapist	1	2	1	2
Impersonal Presentation	1	2	1	2
Other	1	2	1	2

Comments

Appendix 2

THERAPIST STRATEGY RATING FORM

Patient:

Therapist:

Therapy Date:

Rater:

Consultation Date:

Therapy Session:

Interpersonal Focus:

Grief Yes No

Role Transitions Yes No

Role Disputes Yes No

Interpersonal Deficits Yes No

Goal Directed Activity

Quality

	Yes	No	Rec	Excellent						Poor
Exploration of recent and remote losses and reactions to these losses	1	2	3	1	2	3	4	5	6	7
Facilitation of Mourning	1	2	3	1	2	3	4	5	6	7
Exploration of ways patient can develop and/or resume relationships and activities	1	2	3	1	2	3	4	5	6	7
Information gathering and exploration re: nature of disputes and/or role transition	1	2	3	1	2	3	4	5	6	7
Clarification of the patient's position in the disputes and/or transition	1	2	3	1	2	3	4	5	6	7
Exploration and discussion of possible changes that could be made	1	2	3	1	2	3	4	5	6	7
Review of past and current relationships in detail	1	2	3	1	2	3	4	5	6	7
Review of self concept, with emphasis on self-destructive, unrealistic attitudes/expectations	1	2	3	1	2	3	4	5	6	7
Careful attention to the positive and negative elements of the patient/therapist relationship	1	2	3	1	2	3	4	5	6	7

Comments

Appendix 2

TERMINATION SESSION(S) CHECKLIST AND RATING FORM

Patient:

Therapist:

Therapy Date:

Rater:

Consultation Date:

Therapy Session:

TASKS

QUALITY

	Yes	No	Rec	Excellent				Poor			
Explicit discussion of the end of treatment	1	2	3	1	2	3	4	5	6	7	
Elicit/discuss patient's reaction to termination	1	2	3	1	2	3	4	5	6	7	
Acknowledgement of the end of treatment as a time of potential grieving	1	2	3	1	2	3	4	5	6	7	
Help patient move toward a recognition of his/her independent competence	1	2	3	1	2	3	4	5	6	7	
Review the course of treatment and progress with the patient	1	2	3	1	2	3	4	5	6	7	
Patient given opportunity to evaluate the treatment and assess future needs	1	2	3	1	2	3	4	5	6	7	
Assess with patient early warning signals, and discuss procedures for re-entry into treatment if necessary.	1	2	3	1	2	3	4	5	6	7	

Comments

Appendix 3

Level A: Interest

- Available for health care professionals who are interested in IPT. This will provide an overview in the form of an introductory training course lasting two days or more.

Level B: Basic Training as IPT Therapist

- Trainees should have read the IPT manual and attended a training course of 2-4 days.
- Supervision is offered at the discretion of the supervisor.
- Supervisees should have previous clinical training with a good knowledge of mood disorders.
- First case using IPT should be in the treatment of major depression and ideally should be provided without adjunctive antidepressant medication.
- Each trainee should be supervised for a minimum of 2 cases on model.
- All sessions should be recorded (video/audio) & a minimum of 3 tapes from each case selected randomly by the supervisor for formal review e.g. using the IPT competency scale. A minimum of 12 out of 16 sessions per case will be supervised. Supervision can be individual group, but each trainee should receive at least 4 hours supervision for each case. In group supervision, the trainee will have the opportunity to discuss their own case for 4 hours. The 2 cases should be in 2 different focal areas.
- Satisfactory supervisors report provided when the above criteria is met e.g. 'x has attended an introductory course in IPT and has achieved a satisfactory standard in 2 supervised cases'

Level C: CPD for IPT Therapists

- IPT Therapists should carry an IPT caseload - at least 2 cases a year.
- IPT Therapists should receive on going supervision, at least monthly, this may be individual, peer group or even via the telephone

Level D: IPT Supervisor

- To have achieved level A, B and to continue level C.
- Minimum of 10 supervised cases (at least 2 in each focal area). This supervision may be individual or group, and includes the 2-3 cases in level B.
- Supervisors would be required to attend a Network of supervisors. It is proposed that regional groups be established and meet at least twice a year.
- Supervisors must attend an introductory Supervisor's workshop before providing supervision. It is proposed that workshops will be run twice yearly by existing supervisors. There will move around the country and are likely to run alongside one of the established training courses.
- It is recommended to attend the Annual Meeting of the IPT interest group which will rotate around the UK.
- Supervisors should be prepared to keep their IPT clinical & supervisory skills active by supervising at least 2 trainings per year and keeping level C activity for clinical work.

Appendix 4

Therapist Details

Age: Sex: Male/Female

Profession: Degree(s) Held:

Clinical experience

Clinical Qualification(s):

Years of Clinical Practice:

Number of depression cases treated: 1-10 10-20 30+
(any treatment model)

No of cases treated using the IPT model:

- at time of starting training cases-
- current total-

Preferred Treatment Model(s):
(for depression)

IPT Details

Date of group training:

Date casework began:

Number of supervision sessions
attended during training cases:

Appendix 4

Patient details

	First case	Second case	Third case
Patient's Initials: (please circle)			

Sex:

Age:

Marital Status:

Education:

Employment:

Diagnosis:

Previous Diagnoses: e.g. depression, anxiety

Previous Treatment:

Problems in conducting therapy:

BIBLIOGRAPHY

BIBLIOGRAPHY

- Ablon, J.S. & Jones, E.E. (1999). Psychotherapy Process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, Vol. 67, No.1, 64-75.
- Agras, W. S., Telch C.F., Arnow, B., Eldredge, K., Detzer, M.J., Henderson, J. & Marnell, M. (1995). Does Interpersonal Therapy Help Patients with Binge Eating Disorder Who Fail to Respond to Cognitive-behavioural Therapy. *Journal of Consulting and Clinical Psychology*, Vol. 63, No. 3, 356-360.
- Agras, W.S., Walsh, B.T., Wilson, G.T. & Fairburn, C.G. (1999). A multiste comparison of cognitive-behaviour therapy (CBT) and Interpersonal psychotherapy (IPT) I the treatment of bulimia nervosa. Presented at Eating Disorders '99, London, April 20-22.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition. Washington, DC, American Psychiatric Association.
- Auerbach, A.H. & Johnson, M. (1977). Research on the therapist's level of experience. In Gurman, A.S. & Razin, A.M. (Eds) *Effective Psychotherapy: A Handbook of Research*. New York: Pergamon Press.
- Balastrieri, M., Williams, P. & Wilkinson, G. (1988). Specialist mental health treatment in general practice: A meta analysis. *Psychological Medicine*, 18 (3), 711-717.
- Barber, J.P., Crits-Christoph, P. & Luborsky, L. (1996). Effects of Therapist Adherence and Competence on Patients Outcome in Brief Dynamic Therapy. *Journal of Consulting and Clinical Psychology*, 64 (3), 619-622.
- Barber, J.P. & Muenz, L.R. (1996). The Role of Avoidance and Obsessiveness in Matching Patients to Cognitive and Interpersonal Psychotherapy: Empirical Findings from the Treatment for Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, 64 (5), 951-958.
- Barkham, M. (1989a). Brief prescriptive therapy in two-plus-one sessions: I. Initial cases from the clinic. *Behavioural Psychotherapy*, 17, 161-175.
- Barkham, M. (1989b). Exploratory therapy in two-plus-one sessions: I. Rationale for a brief psychotherapy model. *British Journal of Psychotherapy*, 6, 79-86.
- Barkham, M. & Hobson, R.F. (1989). Exploratory therapy in two-plus-one sessions: II. A single case study. *British Journal of Psychotherapy*, 6, 87-98.
- Barlow, D.H. (1994). Psychological Interventions in the Era of Managed Competition. *Clinical Psychology: Science and Practice*, 1:109-122.

- Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. & Erbaugh, J.(1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*, 4, 561-571.
- Beck, A.T., Rush, A.J., Shaw, B.F., Emery, G. (1979). *Cognitive Therapy for Depression*. New York: Guilford Press.
- Benjamin, L. S. (1974). Structured Analysis of Social Behaviour. *Psychological Review*, Vol. 81, 392-425.
- Bennett, D., Parry, G & Ryle, A. (in press). Development of a Measure of Therapists Competence in Resolving Transference Enactments Which Threaten the Therapeutic Alliance. *Psychotherapy Research*.
- Berman, J.S. & Norton, N.C. (1985). Does Professional Training Make a Therapist More Effective? *Psychological Bulletin*, Vol. 98, No.2, 401-407.
- Bernard, J.M. (1997). Supervisory training : a discrimination model. *Counsellor Education and Supervision*, (USA), 27, 500-9.
- Beutler,L.E. (1995). Common factors and specific effects. *Clinical Psychology*, Spring, 79-82.
- Binder, J.L. & Strupp, H.H. (1997). In *Handbook of Psychotherapy Supervision*. Watkins, C.E. (Ed.). John Wiley & Sons.
- Blackburn, I.M., Bishop, S., Glen, A.I.M., Whalley, L.J., Chritie, J.E. (1981). The efficacy of cognitive therapy in depression: a treatment trial using cognitive therapy and pharmacotherapy, each alone and in combination. *British Journal of Psychiatry*, Vol. 171, 328-334.
- Blackburn I.M., James,I., Milne, D., Baker, C., Garland,A., Standard, S. & Reichelt, K. (1999). The Revised Cognitive Therapy Scale, II: Psychometric properties. Unpublished manuscript.
- Blanco,C., Lipsitz,J. & Caligor, E. (2001). Treatment of Chronic Depression with a Twelve Week Program of Interpersonal Psychotherapy. *American Journal of Psychiatry*, Vol. 158, No. 3, 371-375.
- Blatt, S.J., Quinlan, D.M., Pilkonis, P.A. & Shea, M.T. (1995). Impact of perfectionism and need for approval on the brief treatment of depression: The National Institute of Mental Health Treatment of Depression Collaborative Research Program Revisited. *Journal of Consulting and Clinical Psychology*, 63 (1), 125-132.
- Blatt, S.J., Sanislow III, C.A., Zuroff, D.C. & Pilkonis, P.A. ((1996). Characteristics of Effective Therapists: Further Analyses of Data From the National Institute of Mental Health Treatment of Depression Collaborative Program. *Journal of Consulting and Clinical Psychology*, Vol. 64, No. 6, 1276-1284.
- Boelens, W. (1990). Cognitive and behavioural therapy of depression. Dissertation, University of Groningen, The Netherlands.

- Bowlby, J. (1969). *Attachment*. New York: Basic Books.
- Brown, G.W., Harris, T., & Copeland, J.R. (1977). Depression and loss. *British Journal of Psychiatry*, 130: 1-18.
- Brown, C.H.C., Schulberg, M.J., Madonia, M.K., Shear, M.K. & Houck, P.R. (1996) Treatment outcomes for primary care patients with major depression and lifetime anxiety disorders *American Journal of Psychiatry* 153(10) pp1293-1300
- Brown, C., Schulberg, D.S., Perel, J.M. & Houck, P.R. (1999). Effectiveness of treatments for major depression in primary care: a post hoc analysis of outcomes for African American and white patients. *Journal of Affective Disorders*, Vol 53, 185-192.
- Buckley, P., Conte, H.R., Plutchik, R., Wild, K.V. & Karasu, T.B. (1984). Psychodynamic variables as predictors of psychotherapy outcome. *American Journal of Psychiatry*, 141: 742-748.
- Burlingame, G.M., Fahrman, A., Paul, S. & Ogles, B.M.(1989). Implementing a time limited therapy program: Differential effects of training and experience. *Psychotherapy*, 26 (3), 303-313.
- Butler, S.F., Henry, W.P. & Strupp, H.H. (1992). *Measuring adherence and skill in time limited dynamic psychotherapy*. Unpublished manuscript.
- Carroll, K.M., Nich, C., Sifry, R.L., Nuro, K.F., Frankforter, T.L., Ball, S.A., Fenton, L. & Rounsaville, B.J. (2000). A general system for evaluating therapists adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, Vol. 57, 225-238.
- Chapman, L.J. & Chapman, J.P.(1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 193-204.
- Chevron, E.S. & Rounsaville, B.J. (1983) Evaluating the Clinical Skills of Psychotherapists A comparison of Techniques. *Archives of General Psychiatry*, Vol. 40, Oct., 1129-1132
- Chevron, E.S., Rounsaville, B.J., Rothblum, E.D. & Weissman, M.M. (1983). Selecting Psychotherapists to participate in psychotherapy outcome studies. Relationship between psychotherapist characteristics and Assessment of Clinical Skills. *Journal of Nervous and Mental Disease*, Vol. 171, No. 6, 348-353.
- Connor, M. (1999). Training and Supervision Make a Difference. In C. Feltham (Ed) *Controversies in Psychotherapy and Counselling* (pp174-181). London: Sage Publications.
- Crits-Christoph, P. (1998) The Interpersonal Interior of Psychotherapy. *Psychotherapy Research*, Vol. 8, No. 1, 1-16.

- Crits-Christoph, P. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, Vol. 1, No. 1, 81-91
- Crits-Christoph, P., Chambless, D.L., Frank, E. & Brody, C (1995). Training in empirically validated treatments: What are clinical psychology students learning? *Professional Psychology – Research and Practice*, 26 (5), 514-522.
- Davidson, G.C. & Lazarus, A A. (1995). The dialectics of science and practice. In Hayes, S C., Follette, V.M., Dawes, R.M. & Grady, K.E. (Eds), *Scientific Standards of Psychological Practice: Issues and Recommendations* (pp 95-120). Reno, NV: Context Press.
- Davies-Osterkamp, S., Straus, B.M. & Schmitz, N. (1996). Interpersonal problems as predictors of symptom related treatment outcome. *Psychotherapy Research*, 6, 164-176.
- Dawes, R.M., Faust, D. & Meehl, P.E. (1989) Clinical versus Actuarial Judgement. *Science*, Vol.243, 1668-1674.
- Dawes, R.M. (1994). *Psychology and psychotherapy built on myth*. New York: Free Press.
- Dawson, D.F. (1988). Treatment of the borderline patient, relationship management. *Canadian Journal of Psychiatry*, 33: 370-74.
- DeRubeis, R.J., Hollon, S.D., Evans, M.D. & Bemis, K.M. (1982). Can Psychotherapies for Depression be Discriminated? A systematic Investigation of Cognitive Therapy and Interpersonal Therapy. *Journal of Consulting and Clinical Psychology*, Vol.50, No.5, 744-756.
- DeRubeis, R.J. & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research*, 14, 469-482.
- Dewald, P.A (1997). The process of supervision in psychoanalysis. In *Handbook of Psychotherapy Supervision*. Watkins, C.E. (Ed.). John Wiley & Sons.
- Diamond, J. & Martinovich, Z. (1997). The Impact of training on therapist's use of interventions. *Psychotherapy Research* 7(3) 239-247.
- DiMascio, A., Weissman, M.M., Neu, C., Zwillling, M & Klerman, G.L. (1979). Differential symptom reduction by drugs and psychotherapy in acute depression. *Archives of General Psychiatry*, 36, 1450-1456.
- Dobson. K.S., Shaw, B.F. & Valis, T.M. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology*, 24, 295-300.
- Dobson, K.S. & Shaw, B.F. (1988). The Use of Treatment Manuals in Cognitive Therapy: Experience and Issues. *Journal of Consulting and Clinical Psychology*, Vol.56, No. 5, 673-680.
- Durlak, J.A. (1979). Comparative effectiveness of paraprofessional and professional helpers. *Psychological Bulletin*, 86 (1), 80-92.

- Einhorn, H.J. (1988). Reasoning, Inference and Judgement in Clinical Psychology. In Turk D.C. & Salovey P. (Eds), *Reasoning, Inference and Judgement in Clinical Psychology* (51-70). New York: Free Press.
- Elkin Waskow, I. (1983). Specification of the Technique Variable in the NIMH Treatment of Depression Collaborative Research Program. In *Psychotherapy Research: Where are we and where should we go?* J.B.W. Williams, & R.L. Spitzer, (Eds) Guilford Press, New York, London.
- Elkin, I., Parloff, M.B., Hadley, S.W. & Autry, J.H. (1985). NIMH Treatment of Depression Collaborative Research Program: Background and research plan. *Archives of General Psychiatry*, 42, 305-316.
- Elkin, I., Pilkonis, P.A., Docherty, J.P., & Sotsky, S.M. (1988). Conceptual and methodological issues in comparative studies of psychotherapy and pharmacotherapy, I: Active ingredients and mechanisms of change. *American Journal of Psychiatry*, 145, 909-917.
- Elkin, I., Shea, M.T., Watkins J.T., Imber, S.D., Sotsky, S.M., Collins, J.F., Glass, D.R., Pilkonis, P.A., Leber, W.R., Docherty, J.P., Fiester, S.J. & Parloff, M.B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry*, 46, 971-982.
- Elkin, I (1994). The NIMH Treatment of Depression Collaborative Research Program: Where we began and where we are. In *Handbook of Psychotherapy and Behaviour Change*, A.E. Bergin & S.L. Garfield (Eds.) (4th ed.), 114-139. New York: Wiley.
- Elkin, I., Gibbons, R.D., M.T., Watkins, Shea, M.T., Sotsky, S.M., Pilkonis, P.A., Hedecker, D (1995). Initial Severity and Differential Treatment Outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, Vol. 63, No.5, 841-847.
- Elkin, I., Yamaguchi, J.L., Arnkoff, D.B., Glass, C.R., Sotsky, S.M., Krupnick, J.L. (1999). "Patient-Treatment Fit" and early engagement in therapy. *Psychotherapy Research*, 9 (4), 437-451.
- Elkin, I. (1999). A Major Dilemma in Psychotherapy Outcome Research: Disentangling Therapists from Therapies. *Clinical Psychology: Science and Practice*, Vol 6, 10-32.
- Emmelkamp, P.M.G., Bouman T.K. & Blaauw, E. (1994). Individualised versus standardised therapy: a comparative evaluation with obsessive-compulsive patients. *Clinical Psychology and Psychotherapy*, Vol.1 (2), 95-100.
- Fairburn, C.G., Jones, R., Peveler, R.C., Carr, S.J., Solomon, R.A., O'Connor, M., Burton, J. & Hope, R.A. (1991). Three Psychological Treatments for Bulimia Nervosa. *Archives of General Psychiatry*, 48, 463-469.

- Fairburn, C.G., Jones, R., Peveler, R.C., Hope, R.A. & O'Connor, M.I (1993) Longer term effects of Interpersonal Psychotherapy, Behaviour Therapy and Cognitive Behaviour Therapy. *Archives of General Psychiatry*, 50, June, 419-428.
- Fairburn, C.G., Norman, P.A, Welch, S.L., O'Connor M.E., Doll, H.A. & Peveler, R.C. (1995) A Prospective Study of Outcome in Bulimia Nervosa and the Long-term Effects of Three Psychological Treatments. *Archives of General Psychiatry*, 52, Apr., 304-312.
- Feske, U., Frank, E., Kupfer, D.J., Shear, M.K., Weaver, E. (1998) Anxiety as a predictor of response to interpersonal psychotherapy for recurrent major depression: an exploratory investigation. *Depression & Anxiety*, 8(4), 135-141
- Filak, J., Abeles, N. & Norquest, S. (1986). Clients pretherapy interpersonal attitudes and psychotherapy outcome. *Professional Therapy, research & Practice*, 17 (3), 217-22.
- Foley, S.H., O'Malley, S., Rounsaville, B.J., Prusoff, B.A. & Weissman, M.M. (1987). The relationship of patient difficulty to therapist performance in interpersonal psychotherapy of depression. *Journal of Affective Disorders*, 12, 207-217.
- Frank, E., Kupfer, D.J., Jacob, M., Jarrett, D. (1987). Personality Features and Response to Acute Treatment in Recurrent Depression. *Journal of Personality Disorders*, Vol.1, No.1, 14-26.
- Frank, E. (1991). Interpersonal Psychotherapy as a maintenance therapy for patients with recurrent depression. *Psychotherapy*, 28, 259-266.
- Frank, E. & Spanier, C. (1995). Interpersonal Psychotherapy for Depression: Overview, Clinical Efficacy, and Future Directions. *Clinical Psychology Science and Practice* 2 (4): 349-369.
- Frank, E., Shear, M.K., Rucci, P., Cyranowski, J.M., Endicott, J., Fojioini, A., Grochocinski, V.J., Houck, P., Kupfer, D.J., Maser, J.D. & Cassano, G.B. (2000) Influence of Panic-Agoraphobic spectrum symptoms on treatment response in patients with recurrent major depression *American Journal of Psychiatry*, 157(7), 1101-1107.
- Fruzzetti, A.E., Walt, J.A. & Linehan, M.M. (1997). Supervision in Dialectical Behaviour Therapy. Watkins Jr. (Ed) *Handbook of Psychotherapy Supervision* (pp84-100). New York: John Wiley & Sons Inc.
- Gaffan, E., Tsaousis, I. & Kemp-Wheeler, S. (1995). Researcher Allegiance and Meta-Analysis: The Case of Cognitive Therapy for Depression. *Journal of Consulting and Clinical Psychology*, Vol. 63, No. 6, 966-980.
- Garfield, S. L.(1992). Major Issues in Psychotherapy Research. In D.K. Freedheim, (Ed). *History of Psychotherapy: A Century of Change*

- Goldfried, M.R., Castonguay, L.G., Hayes, A.M., Drozd, J.F., Shapiro, D.A. (1997). A comparative analysis of the therapeutic focus in cognitive-behavioural and psychodynamic-interpersonal sessions. *Journal of Consulting and Clinical Psychology*, Vol. 65, No.5, 740-748.
- Grawe, K. (1997). Research Informed Psychotherapy. *Psychotherapy Research*, 7(1), 1-19.
- Guest, P.D. & Beutler, L.E. (1988). Impact of Psychotherapy supervision on therapist orientation and values. *Journal of Consulting and Clinical Psychology*, 56, 653-658.
- Hamilton, M. (1960). A Rating Scale for Depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56-62.
- Hamilton, M. (1967). Development of a Rating Scale for Primary Depressive Illness. *British Journal of Social and Clinical Psychology*, 6, 278-296.
- Hardy, G.E. & Shapiro, D.A. (1985). Therapist verbal response modes in prescriptive vs. Exploratory psychotherapy. *British Journal of Clinical Psychology*, 24, 235-245.
- Hardy, G.E., Barkham, M., Shapiro, D.A., Reynolds, S. & Rees, A. (1995). Credibility and outcome of cognitive-behavioural and psychodynamic-interpersonal psychotherapy. *British Journal of Clinical Psychology*, 34, 555-569.
- Hattie, J.A., Sharpley, C.F. & Rogers, H.J. (1984). Comparative effectiveness of professional and paraprofessional helpers. *Psychological Bulletin*, 95, 534-541.
- Hawkins, P. & Shoet, R. (1989). *Supervision in the Helping Professions*. Milton Keynes, Philadelphia: Open University Press
- Henry W.P., Strupp, H.H., Butler, S.F., Schacht, T.E. & Binder, J.L. (1993). Effects of training in time-limited dynamic psychotherapy: Change in therapists' behaviour. *Journal of Consulting and Clinical Psychology*, 61, 434-440.
- Henry W.P., Schacht, T.E. Strupp, H.H., Butler, S.F., & Binder, J.L. (1993). Effects of training in time-limited dynamic psychotherapy: Mediators of therapists' responses to training. *Journal of Consulting and Clinical Psychology*, 61, 441-447.
- Herbert, J.D. & Mueser, K.T. (1991). The proof is in the pudding: a commentary on Persons. *American Psychologist*, 46, 1347-1348.
- Hiatt, D & Hargrave, G.E. (1995). The characteristics of highly effective therapists in managed behavioural provider networks. *Behavioural Healthcare Tomorrow*, 4, 19-22.
- Hickling, E.J. & Blanchard, E.B. (1997). The Private Practice Psychologist and Manual-based Treatments: Post-Traumatic Stress Disorder secondary to Motor Vehicle Accidents. *Behavioural Research and Therapy*, Vol.35, No.3, 191-203.

- Hill, C.A., O'Grady, K.E. & Elkin, I. (1992). Applying the Collaborative Study Psychotherapy Rating Scale to Rate therapist adherence in Cognitive-Behavioural Therapy, Interpersonal Therapy, and Clinical Management. *Journal of Consulting and Clinical Psychology*, Vol.60, No.1, 73-79.
- Hobson R.F. (1985). *Forms of Feeling: The heart of psychotherapy*. New York: Basic Books.
- Hogan, R. A. (1964). Issues and Approaches in Supervision. *Psychotherapy: Theory, Research and Practice*, 1, 139-141.
- Hollon, S.D., Evans, M.D., Aurebach, A., DeRubeis, R.J., Elkin, I., Lowery, A., Kriss, M., Grove, W., Tuason, V.B. & Piasecki, J. (1988). *Development of a system of rating therapies for depression: Differentiating cognitive therapy, interpersonal psychotherapy, and clinical management psychotherapy*. Unpublished manuscript.
- Hollon, S.D. & Beck, A.T. (1994). Cognitive and cognitive-behavioural therapies. In A.E. Bergin & S.L. Garfield (Eds.), *Handbook of psychotherapy and behaviour change* (4th ed.) 428-466. New York: Wiley
- Horowitz, M.J., Marmar, C., Weiss, D.S., De Witt, K.N. & Rosenbaum, R. (1984b). Brief Psychotherapy of bereavement reactions. *Archives of General Psychiatry*, Vol. 41, 438-448.
- Horowitz, L.M., Rosenberg, S.E., Baer, B.A. Ureno, G. & Villasenor, V.S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical application. *Journal of Consulting & Clinical Psychology*, 56, 885-892.
- Hovarth, A.O. & Symonds, B.D. (1991). Relation between working alliance and outcome in psychotherapy: A meta analysis. *Journal of Counselling Psychology*, 38, 139-149.
- Imber, S.D., Pilkonis, P.A., Sorsky, S.M., Elkin, I., Watkins, J.T., Collins, J.F., Shea, M.T., & Leber, S.R. (1991). Mode-specific effects among three treatments for depression. *Journal of Consulting and Clinical Psychology*, 58, 352-259.
- Jacobson, N.S., Schmalings, K.B., Holtzworth-Munroe, A., Katt, J.L., Wood, L.F. & Follette, V.M. (1989). Research-Structured V's Clinically Flexible Versions of social-learning based marital therapy. *Behavioural Research and Therapy*, Vol. 27, No.2, 173 -180.
- Jarrett, R.B., Rush, A.J. (1994). Short term psychotherapy of depressive disorders: current status and future directions. *Psychiatry*, 57, 115-132.
- Johnson, S.L., Monroe, S., Simons, A. & Thase, M.E. (1994). Clinical characteristics associated with interpersonal depression: symptoms, course and treatment response. *Journal of Affective Disorders*, Vol.31 97-109.
- Karasu T.B. (1990). Towards a Clinical Model of Psychotherapy for Depression, I: Systematic Comparison of Three Psychotherapies. *The American Journal of Psychiatry* 147: 2, February.

- Karasu T.B. (1990). Towards a Clinical Model of Psychotherapy for Depression, II: An Integrative and Selective Treatment Approach. *The American Journal of Psychiatry* 147: 3, March.
- Kazdin A.E. & Bass, B. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, Vol. 57, No. 1, 138-147.
- Kazdin, A.E. (1994). Methodology, design and evaluation in psychotherapy research. In A.E. Bergin & S.L. Garfield (Eds) *Handbook of psychotherapy and behaviour change* (4th ed.) (.19 71). New York: Wiley.
- Kiesler, D.J. (1994). Standardisation of intervention: The tie that binds psychotherapy research and practice. In P.F. Talley, H.H. Strupp, & S.F. Butler (Eds.) *Psychotherapy research and practice: Bridging the gap* (143-153). New York: Basic Books.
- Kingdon, D., Tyrer, P., Seivewright, N., Ferguson, B. & Murphy, S. (1996). The Nottingham Study of Neurotic Disorder: Influence of Cognitive Therapists on Outcome. *British Journal of Psychiatry*, 169, 93-97.
- Klein, D.F. & Rabkin, J. G. (1985). Specificity and Strategy in Psychotherapy Research and Practice. In *Psychotherapy Research: Where are we and where do we go?* J.B.W. Williams & R.L. Spitzer (Eds.), New York, London: Guilford Press.
- Klein, D.F. (1997). A Psychotherapeutic Context for Clinical Trials Is Promising, But Manualization Is Not. *Archives of General Psychiatry*, 54, Oct., 929-930.
- Klerman, G.L., DiMascio, A., Weissman, M.M., Prusoff, B., & Paykel, E. (1974). Treatment of depression with drugs and psychotherapy. *American Journal of Psychiatry*, 131, 186-191.
- Klerman, G.L. & Weissman, M.M., Rounsaville, B.J., Chevron, E.S. (1984). *Interpersonal Psychotherapy of Depression*. New Jersey: Jason Aronson Inc.
- Klerman, G.L. & Weissman, M.M. (Eds.) (1993). *New Applications of Interpersonal Psychotherapy*. American Psychiatric Press Inc., Washington DC, London.
- Krupnick, J.L. & Pincus, H.A. (1992). The Cost-effectiveness of Psychotherapy: A Plan for research. *American Journal of Psychiatry*, 149: 10, Oct, 1295- 1305.
- Krupnick, J.L., Elkin, I., Collins, J., Simmens, S., Sotsky, S.M., Pilkonis, P.A., & Watkins, J.T. (1994). Therapeutic Alliance and Clinical Outcome in the NIMH Treatment of Depression Collaborative Research Program: Preliminary Findings. *Psychotherapy*, 31, 1, 28-35.
- Krupnick, J.L., Sotsky, S.M., Simmens, S., Moyer, J., Elkin, I., Watkins, J. & Pilkonis, P.A. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome. Findings in the NIMH Treatment of Depression Collaborative Research Program. *journal*

- Kupfer, D.J.E, Frank, E. Perel, M.J. (1989). The advantage of early treatment intervention in recurrent depression. *Archives of General Psychiatry*, 46, 771-775.
- Lafferty, P., Beutler, L.E. & Crago, M. (1989). Differences between more and less effective psychotherapists: A study of Select Therapist Variables. *Journal of Consulting and Clinical Psychology*, Vol. 57, No.1, 76-80.
- Lambert, M.J. & Bergin, A. E. (1992). Achievements and Limitations of Psychotherapy Research. In *History of Psychotherapy: A Century of Change*. D.K. Freedheim, (Ed)
- Lang, P.J., Lazovik, A.D. & Reynolds, D.J. (1965). Desensitization, suggestibility and pseudopsychology. *Journal of Abnormal Psychology*, 70, 395-402.
- Lave, J.R., Frank, R.G., Schulberg H.C. & Kamlet, M.S. (1998). Cost-Effectiveness of Treatments for Major Depression in Primary Care Practice. *Archives of General Psychiatry*, 55:645-651.
- Liese, B.S. Barber, J. & Beck, A.T. (1995). *The Cognitive Therapy Adherence and Competence Scale*. Unpublished instrument. University of Kansas Medical Centre, Kansas City.
- Liese, B.S. & Beck, A.T. (1997). Back to basics: Fundamental Cognitive Therapy skills for keeping drug dependent individuals in treatment. In J.J. Boren, L.S. Onken, & J.D. Blaine (Eds) *Beyond the therapeutic alliance: Keeping drug dependent individuals in treatment*. (210-235). National Institute on Drug Abuse Monograph. Washington D.C.: Government Printing Office.
- Linehan, M.M. (1993a). *Cognitive-behavioural treatment of borderline personality disorder*. New York: Guilford.
- Linehan, M.M. (1993b). *Skills training manual for treating borderline personality disorder*. New York Guilford.
- Lipsitz, J.D. , Fyer A.J., Markowitz, J.C., & Cherry, S. (1999) An open trial of interpersonal psychotherapy for social phobia. *American Journal of Psychiatry*, 101: 141-148
- Luborsky, L., Singer, B. & Luborsky, L. (1975). Comparative Studies of Psychotherapies. Is it true that "Everyone Has Won and All Must Have Prizes"? *Archives of General Psychiatry*, 32, 995-1008.
- Luborsky, L. (1984). *Principles of Psychoanalytic Psychotherapy. A Manual for Supportive-Expressive Treatment*. Basic Books.
- Luborsky, L., DeRubeis. R.J. (1984). The use of psychotherapy treatment manuals: a small revolution in psychotherapy research style. *Clinical Psychology Review*, 1984; 4: 5-14.
- Luborsky, L., McLellan, A.T., Woody, G.E., O'Brien C.P., Auerbach, A. (1985). Therapists success and its determinants. *Archives of General Psychiatry*, 42: 602-611.

- Luborsky, L. Crits-Christoph, P., McLellan, A.T., Woody, G., Piper, W., Lieberman, B., Imber, S. & Pilkonis, P. (1986). Do therapists vary much in their success? Finding from our four outcome studies. *American Journal of Orthopsychiatry*, 56 (4), Oct. 501-512.
- Malan, D.H. (1963). *A study of brief psychotherapy*. New York: Plenum
- Markowitz, J.C. (1993). Psychotherapy of the post-dysthymic patient. *Journal of Psychotherapy Practice and Research*, 2: 157-163.
- Markowitz, J.C., Klerman, G.L., Clougherty, K.F., Speilman, L.A., Jacobsberg, L.B., Fishman, B., Frances, A.J., Kocsis, J.H., & Perry, S.W. (1995). Individual Psychotherapies for Depressed HIV+ patients. *American Journal of Psychiatry*, 152: 1504-1509.
- Markowitz, J.C. (1998)(Ed.). *Interpersonal Psychotherapy*. (1st ed.) Washington DC, London: American Psychiatric Press, Inc.
- Markowitz, J.C. (1998). *Interpersonal Psychotherapy for Dysthymic Disorder*. Washington D.C., London: American Psychiatric Press, Inc.
- Marziali,E. & Munroe-Blum, H. (1994). *Interpersonal group psychotherapy for borderline personality disorder*. New York, Basic Books.
- Matarazzo, R.G., Philips, J.S., Weins, A.N. & Saalow, G. (1965). Learning the art of interviewing: A study of what beginning students do and their pattern of change. *Psychotherapy: Theory, Research and Practice*, 2, 49-60.
- McNeilly, C.L. & Howard, K.I. (1991). The therapeutic procedures inventory: psychometric properties and relationship to phase of treatment. *Journal of Psychotherapy Integration*, 1, 223-234.
- Meehl, P.E. (1960). The cognitive activity of the clinician. *American Psychologist*, 15, 19-27.
- Meyer, A. (1957). *Psychobiology. A Science of Man*. Springfield, Ill. Charles C. Thomas.
- Milne, D.L., Baker, C., Blackburn, I., James, I. & Reichelt, K. (1999). Effectiveness of cognitive therapy training. *Journal of Behaviour Therapy*, Vol. 30, 81-92.
- Moncher, F.J. & Printz, R.J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, Vol.11, pp.247-266.
- Moos, R.H. (1990). In D.L. Milne *Coping Responses Inventory*. Windsor: NFER-Nelson.
- Moreau, D., Mufson, L., Weissman, M.M. & Klerman, G.L. (1991). Interpersonal psychotherapy for adolescent depression: Description of modifications and preliminary application. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30 (4), 642-51.

- Morrison, L. A. & Shapiro, D. A. (1987). Expectancy and outcome in prescriptive vs. Exploratory psychotherapy. *British Journal of Clinical Psychology*, 26, 59-60.
- Mufson, L., Moreau, D., Weissman, M.M., & Klerman, G.L. (1993). *Interpersonal psychotherapy for depressed adolescents*. New York, Guilford Press.
- Mufson, L., Moreau, D., Weissman, M.M., Wickramaratne, J.M. & Samilov, A. (1994). Modifications of interpersonal psychotherapy with depressed adolescents (IPT-A): Phase I and II studies. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33(5), 695-705.
- Mufson, L., Weissman, M.M., Moreau, D. & Garfinkel, R. (1999). Efficacy of interpersonal psychotherapy for depressed adolescents *Archives of General Psychiatry*, 56, 573-579.
- Mulder, R.T. (2002). Personality Pathology and Treatment Outcome in Major Depression: A Review. *American Journal of Psychiatry*, 159 (3), 359-371.
- Muller-Popkes, K. & Hajak, G. (1996). Interpersonelle Psychotherapie zur Behandlung von Patienten mit primärer Insomnie - Vorläufige Daten zur polysomnographischen Makroanalyse. *Wien Med. Wochenschrift* 153: 303-305.
- Neufeldt, S.A. & Nelson, M.L. (1998). Research In Training Clinics: A bridge between science and practice. *Journal of Clinical Psychology*, Vol. 54(3), 315-327.
- Newman, F.L., Kopta, S.M., McGovern, M.P., Howard, K.I. & McNeilly, C.L. (1988). Evaluating trainees relative to their supervisors during the psychology internship. *Journal of Consulting and Clinical Psychology*, 56, 659-665.
- Neitzel, M.T. & Fisher, S.G. (1981). Effectiveness of professional and paraprofessional helpers: A comment on Durlak. *Psychological Bulletin*, 89 (3), 555-565.
- Ogrodniczuk, J.S. & Piper, W.E. (1999) Measuring Therapist Technique in Psychodynamic Psychotherapies. Development and Use of a New Scale. *Journal of Psychotherapy Practice and Research*, Vol. 8, pp.142-154.
- O'Malley, S.S., Suh, C. S., & Strupp, H.H.(1983). The Vanderbilt Psychotherapy process Scale: A report on the scale development and a process-outcome study. *Journal of Consulting and Clinical Psychology*, 51, 581-586.
- O'Malley, S.S., Foley, S.H., Rounsaville, B.J., Watkins, J.T., Sotsky, S.M., Imber, S.D. & Elkin, I. (1988). Therapist competence and patient outcome in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology*, 56, 496-501.
- Orlinsky, D.E. & Howard, K.I. (1986). Process and outcome in psychotherapy. In S.L. Garfield & A.E. Bergin (Eds.) *Handbook of psychotherapy and behaviour change* (Chapter 8). New York: John Wiley & Sons.

- Parloff, M. B., Waskow, I.E., Wolfe, B.E. (1978). Research on therapist variables in relation to process and outcome. In Garfield, S.L. & Bergin, A.E. (Eds) *Handbook of Psychotherapy and Behavioural Change*. New York: John Wiley & Sons.
- Parsons, T. (1951). Illness and the role of the physician: a sociological perspective. *American Journal of Orthopsychiatry*, 21: 452-460.
- Persons, J.B. (1991). Psychotherapy outcome studies do not accurately represent current models of psychotherapy. A proposed remedy. *American Psychologist*, Vol. 46, No.2, 99-106.
- Pilkonis, P.A., Imber, S.D., Lewis, P. & Rubinsky, P. (1984) A comparative outcome study of individual, group and conjoint psychotherapy. *Archives of General Psychiatry*, Vol. 41, 431-437.
- Piper, W.E., de Carufel, F.L., Szkrumelak, N. (1985). Patient predictors of process and outcome in short-term individual psychotherapy. *Journal of Nervous and Mental Disease*, Vol. 173, no. 12, 726-733.
- Piper, W.E., Azim, H.F.A., Joyce, A.S., & McCallum, M. (1991). Transference interpretations, therapeutic alliance and outcome in short term individual psychotherapy. *Archives of General Psychiatry*, Vol. 48(10), 946-953.
- Prusoff, B.A., Weissman, M.M., Klerman, G.L. & Rounsaville, B.J. (1980) Research Diagnostic subtypes of depression. Their role as predictors of differential response to psychotherapy and drug treatment. *Archives of General Psychiatry*, 37, 796-801.
- Reynolds, C.F., Frank, E., Perel, J.M., Imber, S., Cornes, C., Miller, M., Mazumdar, S., Houck, P.R., Dew, M., Stack, J.A., Pollock, B.G., & Kupfer, D.J. (1999). Nortriptyline and Interpersonal Psychotherapy as Maintenance Therapies for Recurrent Depression: A Randomized Controlled Trial in Patients Older than 59 Years. *Journal of the American Medical Association*, Vol. 281 (1), 39-45.
- Ricks, D.F. (1974). Supershrink: Methods of a therapist judged successful on the basis of adult outcomes of adolescent patients. In D.F. Ricks, M.Roff, & A. Thomas (Eds.) *Life History Research in psychopathology*. Minneapolis: University of Minnesota Press.
- Roberts, A.H., Kewman, D.G., Mercier, L., & Howell, M. (1993). The power of nonspecific effects in healing: Implications for psychosocial and biological treatments. *Clinical Psychology Review*, 13, 375-391.
- Robinson, L.A., Berman, J.S. & Neimeyer, R.A. (1990). Psychotherapy for the Treatment of Depression: A Comprehensive Review of Controlled Outcome Research. *Psychological Bulletin* 108 (1) 30-49.
- Roth, A., Fonagy, P. & Parry, G. (1996). *What works for whom? A Critical review of Psychotherapy Research*. New York, NY: Guilford Press.

- Rounsaville, B.J., Weissman, M.M. & Prusoff, B.A. (1981). Psychotherapy with depressed outpatients: Patient and process variables as predictors of outcome. *British Journal of Psychiatry*, 138, 67-74.
- Rounsaville, B.J., Chevron, E.S. & Weissman, M.M. (1983). Specification of Techniques in Interpersonal Psychotherapy. In *Psychotherapy Research: Where are we and where should we go?* J.B.W. Williams, & R.L. Spitzer (Eds) New York, London: Guilford Press.
- Rounsaville, B.J., Chevron, E.S., Weissman, M.M., Prusoff B.A. & Frank, E. (1986). Training Therapists to Perform Interpersonal Psychotherapy in Clinical Trials. *Comprehensive Psychiatry*, Vol.27, No.4, 364-371.
- Rounsaville, B.J., Chevron, E.S., Prusoff, B.A., Elkin, I., Imber, S. Sotsky, S.& Watkins, J. (1987). The relation between specific and general dimensions of the psychotherapy process in Interpersonal Psychotherapy of depression. *Journal of Consulting and Clinical Psychology*, Vol.55, No.3, 379-384, 1987.
- Rounsaville, B.J. O'Malley, Foley, S. & Weissman, M.M. (1988). Role of Manual-Guided Training in the Conduct and Efficacy of Interpersonal Psychotherapy for Depression. *Journal of Consulting and Clinical Psychology*, Vol. 56, No.5, 681-688.
- Rush, A.J. (1993) Clinical Practice Guidelines. Good News, Bad News or No News? *Archives of General Psychiatry*, Vol. 50, June, 483- 490.
- Ryle, A. (1991). *Cognitive Analytic Therapy: Active Participation in Change*. Chichester: Wiley.
- Sanderson, W.C. & Woody, S. (1995). *Manuals for Empirically Validated Treatments*. A project of the Task Force on Psychological Intervention, Division of Clinical Psychology, American Psychological Association. *The Clinical Psychologist*, Vol. 48, No.4, 7-11.
- Seligman, M.E.P. (1995). The effectiveness of psychotherapy. The Consumer Reports Study. *American Psychologist* 50, 965-974.
- Schauenberg, H., Kuda, M., Sammet, I., & Strack, M. (2000). The Influence of Interpersonal Problems and Symptom Severity on the Duration and Outcome of Short-Term Psychodynamic Psychotherapy. *Psychotherapy Research* 10 (2), 13-146.
- Schaffer, (1982). Multidimensional measures of therapist behaviour as predictors of outcome. *Psychological Bulletin*, 92, 670-681.
- Schatch, T.E. (1991). *Formulation-Based Psychotherapy Research: Some Further Considerations*. *American Psychologist*, 46, 1346-1347.
- Schulberg, H.C. & Scott, C.P. (1991). Depression in primary care: treating depression with interpersonal psychotherapy. In C.S. Austad, W.H. Berman (Ed), *Psychotherapy in Managed*

Health Care: The Optimal Use of Time and Resources (pp 153-170), Washington DC, American Psychological Association.

Schulberg, H.C., Scott, C.P., Madonia, M.J., Imber, S.D. (1993). Applications of interpersonal psychotherapy to depression in primary care practice. In G.L.,Klerman, & M.M.,Weissman, (Eds) *New Applications of Interpersonal Psychotherapy I* (pp265-291).Washington DC, American Psychiatric Press.

Schulberg, H.C., Block, M.R., Madonia, M.J., Scott, C.P., Rodriguez, E., Imber, S.D.,Perel, J., Lave, J., Houck, P.R., Coulehan, J.L. (1996). Treating Major Depression in primary care practice: eight month clinical outcomes. *Archives of General Psychiatry*, Vol.53 (10), 913-919.

Schulte, D., Kunzel, R., Pepping, G. & Schulte-Bahrenberg, T. (1992). Tailor-made versus standardised therapy of phobic patients. *Advances in Behaviour Research and Therapy*, 14, 67-92.

Shapiro, D.A. & Shapiro, D. (1982). Meta analysis of comparative outcome studies: A reply to Wilson. *Behavioural Psychotherapy*, 10 (4), 307-10.

Shapiro, D.A. & Firth, J.A. (1987). Prescriptive vs. Exploratory psychotherapy : Outcomes of the Sheffield Psychotherapy Project. *British Journal of Psychiatry*, 151, 790-799.

Shapiro, D. A., Firth-Cozens, J. & Stiles, W.B. (1989). The Question of Therapists' Differential Effectiveness. A Sheffield Psychotherapy Project Addendum. *British Journal of Psychiatry*, Vol.154, pp383-385.

Shapiro, D.A., Barkham, M., Hardy, G.E., & Morrison, L.A. (1990). The Second Sheffield Psychotherapy Project: Rationale, design and preliminary outcome data. *British Journal of Medical Psychology*, 63, 97-108.

Shapiro,D.A., Barkham, M., Rees, A., Hardy, G.E., Reynolds,S., Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioural and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology*, Vol.62, No.3, 522-534.

Shaw, B.F. (1984) Specification of the training and evaluation of cognitive therapists for outcome studies. In J. Williams & R. Spitzer (Eds.) *Psychotherapy Research: Where are we and where should we go?* (173-188). New York: Guilford Press.

Shaw, B.F. & Dobson, K.S. (1988). Competency Judgements in the Training and Evaluation of Psychotherapists. *Journal of Consulting and Clinical Psychology*, Vol. 56, No.5, 666-672.

Shaw, B.F., Olmstead, M., Dobson, K., Sotsky, S.M., Elkin, I., Yamaguchi, J., Vallis, M.T., Lowery, A., Watkins, J.T. & Imber, S.D. (1999). Therapist Competence Ratings in Relation to Clinical Outcome in Cognitive Therapy for Depression. *Journal of Consulting and Clinical Psychology*, 67 (6), 837-846.

- Shea, T., Pilokonis, Beckham, E., P.A., Collins, J.F., Elkin, I., Sotsky, S.M., & Docherty, J.P. (1990). Personality Disorders and Treatment Outcome in the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry*, Vol. 147, 711-718.
- Shea, T., Elkin, I., Imber, S.D., Sotsky, S.M., Watkins, J.T., Collins, J.F., Pilokonis, P.A., Beckham, E., Glass, D.R., Dolan, R.T. & Parloff, M.B. (1992). Course of Depressive Symptoms Over Follow Up. Findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, 49, 782-787.
- Sifneos, P.E. (1979). *Short term dynamic psychotherapy: Evaluation and technique*. New York: Plenum.
- Silfneos, P.E. (1992). *Short-term anxiety provoking psychotherapy: A treatment manual*. New York: Plenum.
- Silberschatz, G., Fretter, P.B. & Curtis, J.T. (1986). How do interpretations influence the process of psychotherapy. *Psychotherapy*, 26, 40-46.
- Silverman, W.K. (1991). Person's description of psychotherapy outcome studies does not accurately represent psychotherapy outcome studies. *American Psychologist*, 46, 1351-1352.
- Smith, M.L. & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Snyder, D.K. & Wills, R.M. (1991). Risks and challenges of long-term psychotherapy outcome research: Reply to Jacobson. *Journal of Consulting and Clinical Psychology*, 59, 146-149.
- Smith, M.L., Glass, G.V. & Miller, T.I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Hospital.
- Snyder, D.K. & Wills, R.M. (1991). Risks and challenges of long term psychotherapy outcome research: Reply to Jacobson. *Journal of Consulting and Clinical Psychology*, 59, 146-149.
- Sotsky, S.M., Glass, D.R., Shea, M.T., Pilkonis, P.A., Collins, J.F., Elkin, I., Watkins, J.T., Imber, S.D., Leber, W.R., Moyer, J. & Oliveri, M.E. (1991). Patient predictors of response to psychotherapy and pharmacotherapy: Findings in the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry*, 148, 997-1008.
- Spinelli, M.G. & Weissman, M.M. (1997). The clinical application of interpersonal psychotherapy for depression during pregnancy. *Primary Psychiatry*, 10, 50-57.
- Stein & Lambert (1984). On the relationship between therapist experience and psychotherapy outcome. *Clinical Psychology Review*, Vol. 4, 1-16.
- Stewart, J.W., Garfinkel, R., Nunes, E.V., Donovan, S. & Klein, D.F. (1998). Atypical features and Treatment response in the National Institute of Mental Health Treatment of Depression

- Collaborative Research Program. *Journal of Clinical Psychopharmacology*, Vol. 18, No.6, 429-434.
- Stiles, W.B., Shapiro, D.A. & Elliot, R. (1986). "Are all psychotherapies equivalent?" *American Psychologist*, 41, 165-180.
- Stiles W.B., Shapiro, D.A. & Firth Cozen J.A. (1989). Therapist differences in the use of verbal response mode, forms and intents. *Psychotherapy*, 26 (3), 314-322.
- Stiles, W.B., Shapiro, D.A. (1989). Abuse of drug metaphor in Psychotherapy process and outcome research. *Clinical Psychology Review*, 9 (4), 521-43.
- Stiles, W.B., Agnew-Davies, R, Hardy, G.E., Barkham, M.& Shapiro, D.A. (1998). Relations of alliance with psychotherapy outcome: findings in the second Sheffield psychotherapy project. *Journal of Consulting and Clinical Psychology*, Vol. 66, No.5, 791-802.
- Stoltenberg, C.D. & Delworth, U. (1987). *Supervising counsellors and therapists*. San Francisco: Jossey-Bass.
- Stoltenberg, C.D. & McNeill, B.W. (1997). Clinical Supervision From a Developmental Perspective : Research and Practice. In C.E. Watkins Jr. (Ed) *Handbook of Psychotherapy Supervision* (pp184-202). New York: John Wiley & Sons Inc.
- Strupp, H.H. & Hadley, S.(1979).Specific versus non-specific factors in psychotherapy. *Archives of General Psychiatry*, 36, 1125-1136.
- Stuart, S. & Cole, V. (1996). Treatment of depression following myocardial infarction with interpersonal psychotherapy. *Annals of Clinical Psychiatry* 8 (4):203-6.
- Sullivan, H.S. (1957). *Clinical Studies in Psychiatry*. New York: Norton.
- Svartberg, M. (1989a). Manualization and competence monitoring of short term anxiety provoking psychotherapy. *Psychotherapy*, 26, 564-571.
- Svartberg, M. (1989b) Monitoring the competence of dynamic psychotherapists: A practical demonstration. *Acta Psychiatrica Scandinavica*, 80, 292-296.
- Svartberg, M. & Stiles, T.(1991). Comparative effects of short term psychodynamic psychotherapy: A meta analysis. *Journal of Consulting & Clinical Psychology*. 59, 704-714.
- Svartberg, M. & Stiles, T. (1992). Predicting patient change from therapist competence and patient-therapist complementarity in short term anxiety provoking psychotherapy. A pilot study. *Journal of Consulting & Clinical Psychology*, 60, 304-307.
- Svartberg, M. & Stiles, T.C. (1994). Therapeutic alliance, therapist competences, and client change in short term anxiety-provoking psychotherapy. *Psychotherapy Research*, 4, 20-33.

Svartberg, M. (1999). Therapist Competence: Its temporal course, temporal stability, and determinants in short-term anxiety-provoking psychotherapy *Journal of Clinical Psychology*, 55 (10), 1313-1319.

Task Force on Promotion and Dissemination of Psychological Procedures (1995). Training in and dissemination of empirically-validated psychosocial treatments: report and recommendations. *The Clinical Psychologist*, 48, 3-24.

Taylor MP. Reynolds CF 3rd. Frank E. Cornes C. Miller MD. Stack JA. Begley AE. Mazumdar S. Dew MA. Kupfer DJ. (1999). Which elderly depressed patients remain well on maintenance interpersonal psychotherapy alone?: report from the Pittsburgh study of maintenance therapies in late-life depression. *Depression & Anxiety*. 10(2):55-60, 1999.

Thase, M.E. & Friedman, E.S. (1999). Is psychotherapy an effective treatment for melancholia and other severe depressive states? *Journal of Affective Disorders*, Vol. 54, 1-19.

Ursano, R.J. & Hales, R.E. (1986). A Review of Brief Individual Psychotherapies. *The American Journal of Psychotherapies* 143:12, Dec. 1507-1517.

Vallis, T.M., Shaw, B.F. & Dobson, K.S. (1986). The cognitive therapy scale: psychometric properties *Journal of Consulting and Clinical Psychology*, 54, 381-385.

Vallis, T.M., Shaw, B.F. & McCabe, S.B. (1988). The relationship between therapist competency in cognitive therapy and general therapy skill. *Journal of Cognitive Psychotherapy: An International Quarterly*, Vol. 2, No. 4, pp237-249.

Vasques, M.J.T. (1992). Psychologist as clinical supervisor: Promoting ethical practice. *Professional Psychology, Research and Practice*, 23, 197-202.

Waddington, L. (1997). Clinical Judgement and case formulation. *British Journal of Clinical Psychology* 36, 309-311.

Wagner, E.F., Frank, E. & Steiner, S.C. (1992). Discriminating Maintenance Treatments for Recurrent Depression: Development and Implementation of a Rating Scale. *Journal of Psychotherapy Practice and Research*, 1 (3),

Waltz, J., Addis, M.E., Koerner, K. & Jacobson, N.S. (1993). Testing the Integrity of a Psychotherapy Protocol: Assessment of Adherence and Competence. *Journal of Consulting and Clinical Psychology*, Vol.61, No.4, 620-630.

Wampold, B.E. (1997). Methodological Problems in Identifying Efficacious Psychotherapies. *Psychotherapy Research* 7 (1), 21-43.

Weissman, M.M., Klerman, G.L., Paykel, E., Prusoff & Hanson, (1974). Treatment effects on the social adjustment of depressed patients. *Archives of General Psychiatry*, 30, 771-778.

- Weissman, M.M. & Bothwell S. (1976). Assessment of Social Adjustment by Patient Self Report. *Archives of General Psychiatry*, 83, 1111-1115.
- Weissman, A.N. & Beck, A.T. (1978). Development and validation of the Dysfunctional Attitudes Scale: A Preliminary Investigation. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada.
- Weissman, M.M., Prusoff, B.A., DiMascio, A., Neu, C., Gohlaney, M., & Klerman, G.L. (1979). The efficacy of drugs and psychotherapy in the treatment of acute depressive episodes *American Journal of Psychiatry*, 136, 555-558.
- Weissman, M.M., Klerman, G.L., Prusoff, B.A., Sholomska, D. & Padian, N. (1981). Depressed outpatients one year after treatment with drugs and/or interpersonal psychotherapy (IPT). *Archives of General Psychiatry*, 38, 51-55.
- Weissman, M.M., Rounsaville, B.J. & Chevron, E. (1982). Training Psychotherapists to Participate in Psychotherapy Outcome Studies. *American Journal of Psychiatry*, 139:11, November, 1442-1446.
- Weissman, M.M. & Markowitz, J.C. (1994). Interpersonal Psychotherapy. Current Status. *Archives General Psychiatry*, 51, Aug, 599-606.
- Weissman, M.M., Markowitz, J.C. & Klerman, G.L. (2000). *Comprehensive Guide to Interpersonal Psychotherapy*. Basic Books.
- Wierzbicki, M. (1993). *Issues in Clinical Psychology*. Boston, MA: Allyn & Bacon.
- Wilfey, D.E. , Agras, W.S., Telch, C.F., Rossiter, E.M., Schneider, J.A., Golomb Cole, A., Sifford, L. & Raeburn, D. (1993). Group Cognitive-Behavioural Therapy and Group Interpersonal Psychotherapy for the Nonpurging Bulimic Individual: A Controlled Comparison. *Journal of Consulting & Clinical Psychology*, Vol. 61, No 2, 296-305.
- Wilfey, D., Frank, M.A., Welch, R., Spurrell, E.B. & Rounsaville, B.J. (1998) Adapting Interpersonal Psychotherapy to a group format (IPT-G) for binge eating disorder: Towards a model for adapting empirically supported treatments. *Psychotherapy Research*, 8:4, 379-391.
- Wilson, G.T. (1995). Empirically Validated Treatments as a Basis for Clinical Practice: Problems and Prospects. In S.C. Hayes, V.M. Follette, R.M. Dawes & K.E. Grady (Eds.) *Scientific Standards of Psychological Practice: Issues and Recommendations*. (1st ed.) (163-196) Reno: Context Press.
- Wilson, G.T. (1996). Manual based treatments: the clinical application of research findings. *Behavioural Research Therapy*, 34: 295 -314.
- Wilson, G.T. (1996). Treatment of Bulimia Nervosa: When CBT Fails. *Behaviour Research & Therapy*, Vol. 34, No.3, 197-212.

Wilson, G.T. (1997). Treatment Manuals in Clinical Practice. *Behavioural Research Therapy*, 35, 205 -210.

Wolfson, L., Miller, M., Houck, P., Ehrenpreis, L., Stack, J.A., Frank, E., Cornes, C., Mazumdar, S., Kupfer, D.J., Reynolds, C.F. (1997). Foci of Interpersonal Psychotherapy (IPT) in depressed elders: clinical and outcome correlates in a combined IPT/Nortriptyline protocol. *Psychotherapy Research* 7(1), 45-55.

Worthington, E.L. Jr. (1987). Changes in supervision as counsellors and supervisors gain experience. *Professional Psychology: Research and Practice*, 18 (3), 189-208.

Young, J & Beck, A.T. (1980). *Cognitive Therapy Scale: Rating Manual*. Unpublished manuscript. Centre for Cognitive Therapy, Philadelphia, PA.

Zlotnick, C., Johnson, S.L., Miller, I.W., Pearlstein, T. & Howard, M. (2001). Postpartum Depression in Women Receiving Public Assistance: Pilot Study of an Interpersonal-Therapy-Oriented group Intervention. *American Journal of Psychiatry*, Vol. 158, No. 4, 638-640.